Journal of Abnormal Psychology

Editor: Howard F. Hunt, *New York State Psychiatric Institute and Columbia University*

Associate Editor: William N. Thetford, *College of Physicians and Surgeons, Columbia University*

Consulting Editors

Theodore X. Barber
Medfield Research Foundation
Harold Basowitz
New York University
Eugene I. Burdock
New York State Psychiatric Institute and Columbia University
John M. Butler
University of Chicago
Arthur C. Carr
New York State Psychiatric Institute and Columbia University
Loren J. Chapman
Southern Illinois University
John H. Cumming
New York State Department of Mental Hygiene
W. Grant Dahlstrom
University of North Carolina
Jerome D. Frank
Johns Hopkins University
Sol L. Garfield
Columbia University
Ann Magaret Garner
Lincoln, Nebraska
Israel Goldiamond
Institute for Behavioral Research
Harrison G. Gough
University of California, Berkeley
Starke R. Hathaway
University of Minnesota
Ralph Heine
University of Michigan
Ernest R. Hilgard
Stanford University
Lawrence C. Kolb
New York State Psychiatric Institute and Columbia University
Sheldon J. Korchin
University of California
William G. Lawlor
Fordham University
Ardie Lubin
U. S. Naval Hospital, San Diego
Salvatore Maddi
University of Chicago
David Rosenthal
National Institute of Mental Health
Alan K. Rosenwald
Illinois State Psychiatric Institute
Saul Rosenzweig
Washington University, St. Louis
Jacques Rutschmann
New York State Psychiatric Institute and Columbia University
Kurt Salzinger
New York State Psychiatric Institute and Brooklyn Polytechnic Institute
Joseph Schachter
New York State Psychiatric Institute and Columbia University
Roy Schafer
Yale University
John Shlien
University of Chicago
Morris I. Stein
New York University
Charles Strother
University of Washington
Harold L. Williams
University of Oklahoma School of Medicine
C. L. Winder
Michigan State University
Joseph Zubin
New York State Psychiatric Institute and Columbia University

Helen Orr
Managing Editor

Mary C. McGarraghy
Technical Editor

Henry S. Salmon
Advertising and Business Manager

Robert D. Gorman
Subscription Manager

VOLUME 71
1966

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Prince and Lemon Sts., Lancaster Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Copyright © 1966 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 71

ARTICLES

PAGE

Acker, Charles W. See Goldstein, Michael J.
Amarel, Marianne, Cheek, Frances E., & Stierhem, Robert J. **Studies in the Sources of Variation in Cloze Scores: I. The Raters** 444
Barber, Theodore Xenophon, & Calverley, David Smith. **Effects on Recall of Hypnotic Induction, Motivational Suggestions, and Suggested Regression: A Methodological and Experimental Analysis** 169
Barber, Theodore Xenophon, & Calverley, David Smith. **Toward a Theory of "Hypnotic" Behavior: Experimental Analyses of Suggested Amnesia** 95
Bentler, P. M. See Jones, Austin.
Bergin, Allen E. **Some Implications of Psychotherapy Research for Therapeutic Practice** 235
Blum, Gerald S. See Minkowich, Abram.
Bowers, Kenneth. **Hypnotic Behavior: The Differentiation of Trance and Demand Characteristic Variables** 42
Broverman, Inge K. See Phillips, Leslie.
Brownfield, E. Dorothy, & Keehn, Jack D. **Operant Eyelid Conditioning in Trisomy-18** 403
Bruning, James L. See Persons, Roy W.
Butterfield, Earl C. See Zigler, Edward.
Calverley, David Smith. See Barber, Theodore Xenophon.
Carlson, William A. See Ward, William D.
Cheek, Frances E. See Amarel, Marianne.
Chotlos, John W. See Goldstein, Gerald.
Coe, William C., & Sarbin, Theodore R. **An Experimental Demonstration of Hypnosis as Role Enactment** 400
Crockett, Joseph T. See Goldstein, Michael J.
Davidson, P. O., Payne, R. W., & Sloane, R. B. **Cortical Inhibition, Drive Level, and Conditioning** 310
De Mille, Richard, & Licht, Leonard. **Failure of an Alternation-Learning Test to Discriminate between Lobotomized and Unlobotomized Schizophrenics** 60
Dohrenwend, Barbara Snell, & Dohrenwend, Bruce P. **Stress Situations, Birth Order, and Psychological Symptoms** 215
Dohrenwend, Bruce P. See Dohrenwend, Barbara Snell.
D'Zurilla, Thomas J. **Persuasion and Praise as Techniques for Modifying Verbal Behavior in a "Real-Life" Group Setting** 369
Efstathiou, Aglaia, & Morant, Ricardo B. **Persistence of the Waterfall Illusion Aftereffect as a Test of Brain Damage** 300
Eisenman, Russell. **Psychopathology and Sociometric Choice** 256
Eriksen, Charles W. See Price, Richard H.
Farberow, Norman L., & McEvoy, Theodore L. **Suicide among Patients with Diagnoses of Anxiety Reaction or Depressive Reaction in General Medical and Surgical Hospitals** 287
Farina, Amerigo, Holland, Charles H., & Ring, Kenneth. **Role of Stigma and Set in Interpersonal Interaction** 421
Fisher, Rhoda Lee. **Failure of the Conceptual Styles Test to Discriminate Normal and Highly Impulsive Children** 429
Fontana, Alan. See Rosenwald, George C.
Foulkes, David, Spear, Paul S., & Symonds, John D. **Individual Differences in Mental Activity at Sleep Onset** 280
Geer, James H. **Effect of Fear Arousal upon Task Performance and Verbal Behavior** 119
Geer, James H. **Fear and Autonomic Arousal** 253
Geer, James H., & Katkin, Edward S. **Treatment of Insomnia Using a Variant of Systematic Desensitization: A Case Report** 161
Goldstein, Gerald, & Chotlos, John W. **Stability of Field Dependence in Chronic Alcoholic Patients** 420
Goldstein, Michael J., Acker, Charles W., Crockett, Joseph T., & Riddle, John J. **Psychophysiological Reactions to Films by Chronic Schizophrenics: I. Effects of Drug Status** 335
Goodenough, Donald R. See Lewis, Helen B.
Gough, Harrison G. **Appraisal of Social Maturity by Means of the CPI** 189
Guller, Irving B. **Stability of Self-Concept in Schizophrenia** 275
Hare, Robert D. **Psychopathy and Choice of Immediate versus Delayed Punishment** . 25
Hartman, A Arthur, & Nicolay, Robert C. **Sexually Deviant Behavior in Expectant Fathers** 232

HILGARD, ERNEST R., & TART, CHARLES T. Responsiveness to Suggestions Following Waking and Imagination Instructions and Following Induction of Hypnosis 196
HOLLAND, CHARLES H. See FARINA, AMERIGO.
HOWES, DAVIS. Comments on the So-Called Acquiescence Set in Aphasia 151
JACKSON, C. WESLEY, JR., & POLLARD, JOHN C. Some Nondeprivation Variables which Influence the "Effects" of Experimental Sensory Deprivation 383
JACOBSON, GEORGE R. Effect of Brief Sensory Deprivation on Field Dependence 115
JONES, AUSTIN, BENTLER, P. M., & PETRY, GEORGIA. The Reduction of Uncertainty Concerning Future Pain 87
KAGAN, JEROME. Reflection-Impulsivity: The Generality and Dynamics of Conceptual Tempo 17
KAPLAN, FRANCES. Effects of Anxiety and Defense in a Therapy-Like Situation 449
KATKIN, EDWARD S. The Relationship between a Measure of Transitory Anxiety and Spontaneous Autonomic Activity 142
KATKIN, EDWARD S., SASMOR, DOROTHY BLUM, & TAN, RAMON. Conformity and Achievement-Related Characteristics of Depressed Patients 407
KATKIN, EDWARD S. See GEER, JAMES H.
KEEHN, JACK D. See BROWNFIELD, E. DOROTHY.
KENNEDY, TRUDY. See RINGUETTE, EUGENE L.
LERNER, BARBARA. Rorschach Movement and Dreams: A Validation Study Using Drug-Induced Dream-Deprivation 75
LEVIN, RACHEL B. An Empirical Test of the Female Castration Complex 181
LEWIS, HELEN B., GOODENOUGH, DONALD R., SHAPIRO, ARTHUR, & SLESER, IRVING. Individual Differences in Dream Recall 52
LICHT, LEONARD. See DE MILLE, RICHARD.
LOEVINGER, JANE. Three Principles for a Psychoanalytic Psychology 432
MCEVOY, THEODORE L. See FARBEROW, NORMAN L.
MEICHENBAUM, DONALD H. Effects of Social Reinforcement on the Level of Abstraction in Schizophrenics 354
MENDELSOHN, GERALD A. See ROSENWALD, GEORGE C.
MINKOWICH, ABRAM, WEINGARTEN, LINDA L., & BLUM, GERALD S. Empirical Contributions to a Theory of Ambivalence 30
MORANT, RICARDO B. See EFSTATHIOU, AGLAIA.
NICOLAY, ROBERT C. See HARTMAN, A. ARTHUR.
NOVICK, JACK. Symptomatic Treatment of Acquired and Persistent Enuresis 363
PAUL, GORDON L., & SHANNON, DONALD T. Treatment of Anxiety through Systematic Desensitization in Therapy Groups 124
PAYNE, R. W. See DAVIDSON, P. O.
PERSONS, ROY W., & BRUNING, JAMES L. Instrumental Learning with Sociopaths: A Test of Clinical Theory 165
PETRY, GEORGIA. See JONES, AUSTIN.
PHILLIPS, LESLIE, BROVERMAN, INGE K., & ZIGLER, EDWARD. Social Competence and Psychiatric Diagnosis 209
POLLACK, DONALD. Coping and Avoidance in Inebriated Alcoholics and Normals 417
POLLARD, JOHN C. See JACKSON, C. WESLEY, JR.
PORTZ, ALEXIS T. See ROSENWALD, GEORGE C.
PRICE, RICHARD H., & ERIKSEN, CHARLES W. Size Constancy in Schizophrenia: A Reanalysis 155
REITZ, W. E. See WILLNER, A. E.
RIDDLE, JOHN J. See GOLDSTEIN, MICHAEL J.
RING, KENNETH. See FARINA, AMERIGO.
RINGUETTE, EUGENE L., & KENNEDY, TRUDY. An Experimental Study of the Double Bind Hypothesis 136
ROBACK, HOWARD B., & WEBERSINN, ALFRED L. Size of Figure Drawings of Depressed Psychiatric Patients 416
ROSENWALD, GEORGE C., MENDELSHON, GERALD A., FONTANA, ALAN, & PORTZ, ALEXIS T. An Action Test of Hypotheses Concerning the Anal Personality 304
SARBIN, THEODORE R. See COE, WILLIAM C.
SARETSKY, THEODORE. Effects of Chlorpromazine on Primary-Process Thought Manifestations 247
SASMOR, DOROTHY BLUM. See KATKIN, EDWARD S.
SATZ, PAUL. Specific and Nonspecific Effects of Brain Lesions in Man 65
SCHOPLER, ERIC. Visual versus Tactual Receptor Preference in Normal and Schizophrenic Children 108
SCHUTTE, W. See ZUBEK, JOHN P.
SEIDEN, RICHARD H. Campus Tragedy: A Study of Student Suicide 389
SHANNON, DONALD T. See PAUL, GORDON L.
SHAPIRO, ARTHUR. See LEWIS, HELEN B.
SHARAN (SINGER), SHLOMO. Family Interaction with Schizophrenics and Their Siblings 345

Siegman, Aron Wolfe. Father Absence During Early Childhood and Antisocial Behavior 71

Silber, Leon D. See Truax, Charles B.

Sleser, Irving. See Lewis, Helen B.

Sloane, R. B. See Davidson, P. O.

Spain, Bernadette. Eyelid Conditioning and Arousal in Schizophrenic and Normal Subjects 260

Spear, Paul S. See Foulkes, David.

Spohn, Herbert E., & Wolk, William P. Social Participation in Homogeneous and Heterogeneous Groups of Chronic Schizophrenics 147

Stierhem, Robert J. See Amarel, Marianne.

Symonds, John D. See Foulkes, David.

Tan, Ramon. See Katkin, Edward S.

Tart, Charles T. Types of Hypnotic Dreams and Their Relation to Hypnotic Depth . 377

Tart, Charles T. See Hilgard, Ernest R.

Truax, Charles B. Reinforcement and Nonreinforcement in Rogerian Psychotherapy 1

Truax, Charles B., Wargo, Donald G., & Silber, Leon D. Effects of Group Psychotherapy with High Accurate Empathy and Nonpossessive Warmth upon Female Institutionalized Delinquents 267

Ward, William D., & Carlson, William A. Autonomic Responsivity to Variable Input Rates among Schizophrenics Classified on the Process-Reactive Dimensions 10

Wargo, Donald G. See Truax, Charles B.

Webersinn, Alfred L. See Roback, Howard B.

Weingarten, Linda L. See Minkowich, Abram.

Willner, A. E., & Reitz, W. E. Association, Abstraction, and the Conceptual Organization of Recall: Implications for Clinical Tests 315

Wolk, William P. See Spohn, Herbert E.

Zigler, Edward, & Butterfield, Earl C. Rigidity in the Retarded: A Further Test of the Lewin-Kounin Formulation 224

Zigler, Edward. See Phillips, Leslie.

Zubek, John P., & Schutte, W. Urinary Excretion of Adrenaline and Noradrenaline during Prolonged Perceptual Deprivation 328

Volume 71 February 1966 Number 1

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Chicago

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute

SEYMOUR SARASON
Yale University

JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR	FRANCES L. BREWER	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Editorial Assistant	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.

Copyright © 1966 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 1–9

REINFORCEMENT AND NONREINFORCEMENT IN ROGERIAN PSYCHOTHERAPY[1]

CHARLES B. TRUAX

University of Arkansas and Arkansas Rehabilitation Research and Training Center

Excerpts from tape recordings of a single, long-term, successful therapy case handled by Rogers were analyzed to evaluate the adequacy of the client-centered view that empathy, warmth, and directiveness are offered throughout therapy in a manner not contingent upon the patient's behavior. Findings indicate that the therapists respond in a significantly differential way to 5 of the 9 patient behavior classes studied. Concomitantly, significant increases in the emission rates of 4 of the 5 behavior classes were noted throughout therapy. Findings thus indicated significant reinforcement effects in the client-centered therapy.

The present study is aimed at exploring the possibility that important reinforcement effects occur within the transactions of nondirective therapy.

Client-centered theorists have specified the "therapeutic conditions" of empathic understanding and acceptance or unconditional positive regard as two main antecedents to constructive behavioral or personality change in the client (Dymond, 1949; Hobbs, 1962; Jourard, 1959; Rogers, 1951, 1957; Rogers & Truax, 1965; Truax, 1961; Truax & Carkhuff, 1963). Rogers, as the leading exponent of this viewpoint, holds that these "conditions" are primarily attitudinal in nature and are offered in a nonselective fashion to the patient: they are specifically not contingent upon the patients' verbalizations or behaviors. This viewpoint, in pure form, is incompatible with the behavioristic view of therapy and was one basis for the Rogers-Skinner debates (1956).

The basic difference between the views exemplified by Rogers and Skinner is that the latter holds that an effective therapist attempts to alter the patient's behavior while Rogers holds otherwise. Differential reinforcement is one of the procedures used in operant research *positions*. Thus, whether or not Rogers as a therapist uses differential reinforcement, thereby altering patient behavior, is a central question in the basic issue of control which philosophically differentiates the two positions.

The growing body of evidence indicates that the therapist's accurate empathy and unconditional positive regard are significant antecedents to therapeutic change (Rogers, 1962; Rogers, Kiesler, Gendlin & Truax, 1965). This evidence has been used both as support of Rogers's view and as an argument against the behavioristic views of psychotherapy typified by such theorists as Krasner (1962), Wolpe (1958), Eysenck (1952, 1960), and Bandura (1961). The evidence does suggest that when patients receive high levels of empathy and warmth there is significantly more constructive personality and behavioral change then when the patients receive relatively lower levels (Barrett-Lennard, 1962; Bergin & Solomon, 1963; Cartwright & Lerner, 1963; Dickenson & Truax, 1965; Halkides, 1958; Lesser, 1961; Rogers, 1962; Strupp, 1960; Truax, 1961a, 1961b, 1963; Truax & Carkhuff, 1964; Truax, Carkhuff & Kodman, in press; Truax, Wargo, & Silber, 1965; Wargo, 1962; and Whitehorn & Betz, 1954). None of the research just cited, however, *necessarily* argues against a behavioristic view of psychotherapy.

---

[1] Appreciation is gratefully extended to Carl R. Rogers for his freely given consent to the use of the completed successful counseling case recorded at the University of Chicago Counseling Center in 1955. This particular case is perhaps of special significance since it was heavily used by Rogers and others in the development of the "process conception of psychotherapy" and the "Process Scale" developed in 1957. Thanks are also due to James C. Baxter and Leon D. Silber for their critical comments. This work was supported in part by a grant from the Vocational Rehabilitation Administration, No. RD-906-PM.

If, in contrast to Rogers's contention, the therapist does respond differentially to different patient behaviors (i.e., more accepting of and empathic to, some patient behaviors but less accepting of and more directive in response to other patient behaviors) then a reinforcement view would not be inconsistent with the findings. It could be argued that if empathic understanding, warmth (and nondirectiveness) are therapeutic, then it may also be argued that these therapeutic conditions are reinforcing, rewarding, or somehow encouraging, and that the types of patient behavior (presumably more adaptive ones) that are followed by high levels of these therapeutic conditions will consequently increase during the course of therapy. For example, it may be that the "high conditions" therapist offers more intense levels of accurate empathy and unconditional warmth or acceptance on both a nonselective random basis at, say, a 40% rate of reinforcement for all behaviors and, say, an 85% rate for exploration of material relevant to the private self. By contrast the "low conditions" therapist may offer less intense levels of empathy and warmth, with only a 20% rate of reinforcement for all behavior emitted and only a 40% rate of reinforcement for the patient's explorations of private material.

Support for the position exemplified by Rogers, viewed from the findings on empathy and warmth, rests upon the assumption that the therapist offers levels of conditions that do not systematically covary with the verbalizations or behavior emitted by the patient. If this were true (if, say, the level of therapist empathy or warmth did not systematically covary with patient response classes) then differential reinforcement could not account for the research findings of relationships between therapist behavior and patient outcome. On the other hand, if the therapist, in this case Rogers, does systematically vary his level of warmth or his level of empathy depending on the behavior, then Rogers's position would not be supported.

In an attempt to add clarity to this theoretic controversy, an exploratory analysis of a single successful case handled by Rogers was aimed at determining whether or not important reinforcing effects are imbedded in the transactions of client-centered therapy.

Three qualities of the therapist's behavior were studied as potential reinforcers: (*a*) empathic understanding, (*b*) acceptance or unconditional positive regard, and (*c*) directiveness (a negative reinforcer). These therapist behaviors were examined in relation to nine classes of patient behavior in order to determine the presence or absence of differential therapist responding and any consequent changes in the patient behaviors.[2] The patient behaviors studied which might theoretically be of significance were: (*a*) degree of discrimination learnings by the patient, (*b*) ambiguity of patient's statements, (*c*) degree of insight development by the patient, (*d*) degree of similarity of patient's style of expression to that of the therapist, (*e*) problem orientation of the patient, (*f*) degree of patient catharsis, (*g*) degree of patient blocking, (*h*) degree of patient anxiety, and (*i*) degree of patient negative versus positive feeling expression.

CASE ANALYSIS PROCEDURE

Five clinical psychologists rated an unbiased sample of 40 typewritten interaction units consisting of (*a*) a therapist statement, (*b*) a succeeding patient statement, (*c*) the succeeding therapist statement. These interaction units (TPT, Therapist-Patient-Therapist) were designated by code numbers prior to the ratings, and were then assigned in random order to the five clinical psychologists who served as judges. Each judge rated separately each of the nine patient scales and the three therapist scales in different order, so as to minimize rating biases. The ratings were then decoded, and the ratings of the three classes of "reinforcers" were simply correlated separately with the nine classes of patient behavior under examination. The presence of significant correlations would then be positive evidence to indicate systematic, nonrandom use of these reinforcers with particular classes of patient behavior. Thus the question became, for example, "Does the therapist's degree of acceptance significantly covary with the patient's degree of discrimination learning?" If a positive correlation was found, this would indicate that the therapist systematically was most accepting and unconditionally warm when the patient was engaged in discrimination learning, and was least accepting and warm when the patient engaged in very little discrimination learning.

---

[2] Thanks are due to Israel Goldiamond for critical and helpful questions which served as the stimulus for the analysis of change in patient behaviors over time.

The interaction unit sample. The TPT interaction units were selected from the following interviews out of a total of 85 therapy sessions for the complete case, 1, 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, and 85. Two intersection units were taken from each of the above 20 interviews for a total of 40 interaction units. Interviews from which the samples were drawn, with the exception of Numbers 3 and 7, which were added to give more weight to the earlier stages of therapy, were evenly spaced and should constitute an unbiased sample of interviews throughout the therapy case. The two interaction units from each interview were obtained by starting the playback of the recordings at approximately the end of the first and second one-third of the hour-long tape and then listening until the therapist made a statement. Transcriptions started at the therapist's first words and included the ensuing TPT interaction unit. As a result of this sampling procedure, the length of the therapist and the patient statements varied considerably. When measured to the nearest one-tenth of an 80-character type line the range was from 0.4 to 14.0 lines.

The clinical psychologists as raters. Of the five raters, none was trained in client-centered psychotherapy. One was trained in analytic therapy, while the remaining four clinicians described themselves as eclectic in orientation. All five judges had a minimum of 100 hours of supervised training as therapists, and 1,000 hours of experience as therapists. All postdoctoral clinical psychologists, the judges did not know the hypothesis being investigated. Further, they wrongly assumed that some of the more "nondirective" and "directive" statements could not have come from the same therapist.

The rating scales. A set of graphic rating scales were prepared for each of the three therapist behaviors and nine patient behaviors, each having a brief statement of the variable to be rated and horizontal lines 170 millimeters long on which the rating was marked. The rating was made by simply placing an X along the line labeled "most," "very much," to "least," and "very little." The scales for each therapist and patient behavior class stated a bipolarity of the variable. The bipolar definitions for the three reinforcer variables were taken from Rogers's (1951) theoretic descriptions.

Although the actual therapy samples did not contain extreme examples of the scale values, examples from the present case should add meaning to the present study. Two examples of TPT samples, rated high on each of the three reinforcer variables, are presented below.

Therapist acceptance of unconditional positive regard. The first example illustrates relatively high acceptance of the patient as he is feeling disappointed in the early results of therapy. This sample was rated relatively low on both empathy and directiveness:

T: There is nothing that I can do, but I can and do hope that the person will "lift" or something.

P: Yeh, but, well, for something to happen, or things to change, or me to change . . . inside, or . . . things to change outside, so I can change inside or something. Talking about it . . . doesn't really . . . seem to help, this kind of feeling. I mean, well . . . well, I don't know what I mean. I mean, I guess we're just talking about it. Well I don't know what I'm doing (he has been thumping on something—beating on the desk?—long, long pause) . . . I feel so tremendously self-conscious. I don't know, maybe it was the therapy session yesterday, and the other thing that happened yesterday, that has just thrown, so much on me, myself again. And dangerous, I don't know, not dangerous but . . . I feel so un-free today. So . . . Hmm . . . (long pause) . . . Yeh, boy I really am self-conscious . . . sure tensed up inside.

T: I can't get the feel of what you mean by "self-conscious" It's very much aware of yourself? Or something, or generally embarrassed?

The second example illustrates relatively high acceptance of the patient's feelings of inadequacy and dependency. This sample was rated slightly above average on empathy but below average on directiveness:

T: I guess you're saying "I just can't trust those weak, and helpless and inadequate parts of me. I have to have someone to . . ."

P: To really be me. (T: Mmm, mmm) Someone else, you know . . . that's so absurd . . . that would never work. It's the same thing as, as this, uh . . . being afraid of people. It ties in with being afraid. It's like . . . well, you can use any one of a number of examples. If you really want to be someone genuinely . . . or express something genuinely . . . then, all you have to do is feel the slightest tinge of fear and you won't be able to—really. And it's like that with myself . . . It's kind of . . . when I am myself, it kind of echoes on me and makes me afraid. I suddenly hear myself saying that, and then know, "careful" (T: Mmm, mmm) "Hold on here! Lookout!" (T: Mmm, mmm) . . . like that. (T: Mmm, mmm) "You won't be allowed to live if you do that." (T: Mmm, mmm) "You won't be allowed to . . . *anything*" . . . just, "You'll be blown to smithereens if you try that kind of thing."

T: Mmm, So that if you sense yourself . . . being yourself . . . then my (P: I become afraid) Gosh! Lookout! You don't know what you're getting into—you'll be destroyed.

Therapist directiveness. The first example shows the therapist making a direct request to change the topic of discussion. This sample was rated slightly below average in empathy and low in acceptance:

T: Let's talk about something closer to you than that.

P: Or closer to you. I don't understand this at all, because I was really looking forward to

> this all the time, and now I just don't feel very good . . . about having harmed you.
>
> T: You anticipated coming in, and now . . . today.

A second example of directiveness involves a more subtle "leading" of the patient. This sample was rated as average in empathy but above average in acceptance or unconditional positive regard:

> T: It frightens you to even start to put it into words.
>
> P: I guess I'll have to find it with someone else . . . first.
>
> T: You feel that what would be demanded would be . . . put it in terms of "me" and, "you" . . . uh . . . make this the sort of thing you can sort of dimly visualize. I would need to want to really relate to that fine part of you, and find that so personally rewarding that, that in an attempt I would just . . . keep after it, or something. (P interjects: Yeh) One, one phrase that I . . . I'm bringing in my feelings rather than yours, but . . . ever read the poem "The Hound of Heaven"? It's kinda a weird thing, but, uh, the kind of persistent love of God is the whole theme, that, that won't let the person go . . . and, and, I think that's sort of what you're talking about.

Therapist empathy. The first example illustrates an excerpt in which the therapist attempts to verbalize what he senses is the client's uncertainty; this sample received an average rating on acceptance and a slightly above average rating on directiveness:

> T: I've been trying to soak up that tone, uh, I'm not sure I'm right, but does it have some meaning like this, "What is it you want with me? I'm possibly willing to, to meet that, but I don't know what you want." Does that kinda describe it?
>
> P: Yeh, I'm sympathetic, I'll try and do what I can. "Don't be this, and this, and this way to me." What is it? Yeh, that's it.
>
> T: "So if you want me to get in with whatever it is you expect of me, just let me know."

The second example involves a moment when the therapist attempts to reflect the client's feelings and move one step beyond. This sample was rated average on directiveness and acceptance:

> T: Seems as though all the dark things—hurting, and being hurt—and . . . decay, and corruption, ugliness, uhmmm, Death. It's all of those that (P: frightening) that you're afraid of.
>
> P: Yeh . . . stink and corruption and . . . pus, and. . . . There's just as. . . . It's something dark that ties them all together (T: Mmm, uhuh). Something putrid and (T: Mmm, mmm) . . . there are 10 times the words (T: Mmm, mmm) for it . . . (laughs) it scares me.
>
> T: Just to wander into that field verbally, and . . . and even name all these things that have to do with it . . . this dark side of hurting and rottenness . . . that's hurting in itself.

The patient scales measuring the degree of insight developed, the degree of similarity of the patient's style of expression to that of the therapist, the degree of problem orientation, the degree of catharsis, the degree of blocking in thought and feeling, the degree of anxiety present, and the degree of positive-versus negative-feeling expression were defined by the trained clinical psychologists who served as judges. Degree of ambiguity of the patient's statement was defined in terms of its clarity of meaning. The judges were asked to disregard speech disturbances and length of statement in rating ambiguity. Discrimination learning was defined as making new distinctions between old feelings or experiences, and thus included both cognitive and emotional discrimination learning.[3]

FINDINGS AND DISCUSSION

Qualitative Aspects

There are three qualitative aspects exemplified in this case which perhaps are worth noting. The first concerns the style of expression by the therapist: it was characteristic of the therapist to express, restate, or interpret what the patient has been saying by "quoting" what the patient *might well have said* in the first person singular—"In a sense I feel . . . ." Out of the 40 sampled interaction units, 23 involved first person singular quotes while an additional five (for a total of 28 out of 40) involved impersonal quotes of the type: "In a sense it's like feeling . . . ."

A second characteristic of this particular case was the almost total absence of psychological jargon. Few even semitechnical terms such as "anxious" or "hostile" were used by the therapist. Instead, the therapist relied heavily on everyday language that conveys effect. Thus instead of saying "depressed" the therapist says "hopeless badness." The third qualitative characteristic of this case is the tentative character of therapist statements. There is almost universal use of such prefacing remarks as "in a sense," "I guess," and "maybe." This tentative approach might tend to elicit less resistance from the patient so that actual confrontation might sound much like an attempt to agree with the patient.

The Question of Selective Responding

The reliability of each scale, which is given in parentheses under the scale label in Table 1, was estimated by the variance formula presented by Ebel (1951) for the intraclass correlation. As can be seen in Table 1, reliabilities range from .26 to .64 for the classes of patient behavior, and from .48 to .68 for levels of "reinforcement" offered by the therapist.

---

[3] Available from the author.

The low reliabilities obtained on certain classes of patient behavior would make it difficult to detect any but the strongest of relationships. For the present hypothesis of selective reinforcement the absence of particular relationships is not critical. Rather, the *presence* of selective responding (as indicated by some significant relationship between therapist and patient classes of behavior) would be evidence in support of the hypothesis.

The obtained average intercorrelations between the levels of therapist reinforcements and the levels of the selected patient behaviors are presented in Table 1. These average intercorrelations were obtained in the following manner. First a matrix of intercorrelations was generated for each of the five raters separately. The matrices were then inspected separately for correlations which were significant at or beyond the .05 level of significance. Average correlations for the five raters combined were then obtained for those intercorrelations that were significant in three out of five individual rater matrices. All other correlations were recorded as nonsignificant in the present study so that the reported correlations tend to minimize rather than maximize the possiblity of obtaining significant relationships.

The significant intercorrelations presented in Table 1 show a quite different pattern than would be expected if therapist responses were not highly selective in client-centered psychotherapy. If there was no systematic selective use of empathy, acceptance, or directiveness, then all correlations would be nonsignificant and would approach zero. Such is not the case. The therapist significantly tended to respond selectively with differential levels of empathy, warmth, or directiveness to high and low levels of the following classes of patient behavior: (*a*) learning of discriminations about self and feelings, (*b*) a lack of patient ambiguity (patient clarity), (*c*) patient expressions of insight, (*d*) patient verbal expressions that were similar in style to the therapist's way of expressing himself, and (*e*) problem orientation of the patient. Thus, when the patient expressed himself in a style similar to that of the therapist, the therapist was more empathic, more warm and accepting, and less directive. When the patient expressed himself in a style quite different from that of the therapist, the therapist tended to show significantly less empathy, less acceptance or warmth, and more directiveness.

TABLE 1

INTERRELATIONSHIPS BETWEEN THE LEVEL OF THERAPIST REINFORCEMENT AND LEVELS OF PATIENT BEHAVIORS

Classes of patient behavior	Reinforcers		
	Therapist empathy ($r = .48$)	Therapist acceptance UPR ($r = .59$)	Therapist directiveness ($r = .68$)
Patient learning of discriminations ($r = .59$)	.47	.37	*ns*
Patient ambiguity ($r = .35$)	−.35	−.38	.33
Patient insight ($r = .32$)	.46	.37	*ns*
Similarity of patient style of expression to that of the therapist ($r = .57$)	.48	.32	−.31
Problem orientation ($r = .64$)	*ns*	.35	*ns*
Catharsis ($r = .44$)	*ns*	*ns*	*ns*
Blocking ($r = .54$)	*ns*	*ns*	*ns*
Anxiety ($r = .26$)	*ns*	*ns*	*ns*
Patient negative feeling expression ($r = .29$)	*ns*	*ns*	*ns*

No significant relationships were obtained between the therapist's use of empathy, acceptance, or directiveness, and patient behaviors described as blocking, anxiety, negative- versus positive-feeling expression, or catharsis. While it may be that the absence of these relationships might, in part, be accounted for by the relatively low reliabilities of measurement, it also seems likely that Rogers as a therapist does not tend to respond differentially to these classes of patient behavior. In particular, as a theoretician and therapist, Rogers (1957, 1961) has felt it important for the therapist *not* to respond selectively to negative- versus positive-feeling expression.

TABLE 2

ANALYSIS OF CHANGES OVER TIME IN PATIENT RESPONSE CLASSES

Patient response classes	Highest single correlation with therapist "reinforcer"	Grant orthogonal polynomial analysis of variance for trend			*t* Test between first and all later blocks
		F Linear Trend	F Quadratic Trend	F Cubic Trend	
Similarity of patient style of expression to that of the therapist	.48	7.89***	1.20	.85	2.84***
Patient learning of discriminations	.47	3.10	.79	1.05	2.94***
Patient insight	.46	4.73**	1.70	0.75	2.73***
Patient ambiguity	−.38	3.04	1.50	0.91	1.35
Problem orientation	.35	3.28*	1.61	2.10	1.76**
Catharsis	*ns*	6.10**	2.13	1.20	2.03**
Blocking	*ns*	1.50	6.01**	1.50	1.29
Anxiety	*ns*	2.00	0.98	1.70	0.93
Patient negative-feeling expression	*ns*	1.17	0.65	0.89	0.75

* $p \leq .07$ for 1/39 *df* for trend.
** $p \leq .05$ for 1/39 *df* for trend or for 38 *df* for *t*.
*** $p \leq .01$ for 1/39 *df* for trend or for 38 *df* for *t*.

The Further Question of Reinforcement

The above findings are consistent with, but not direct evidence for, the view that the therapist, in this case Rogers, is consciously or unconsciously using empathy, acceptance, and directiveness as reinforcers. The basic property of a reinforcer is that its use with specific classes of behavior leads to consequent changes in the probability of occurrence of these classes of behavior.

From Table 1, the nine classes of patient behavior can be ranked according to the degree of contingency between therapist "reinforcer" responses and patient responses. Now, if the therapist's systematic selective responding has the properties of reinforcement it would be predicted that, other things being equal, the five patient classes of behavior that were selectively "reinforced" would show increases over time in therapy, while the four classes of patient behavior not reinforced would show no such increase over time. Thus, for example, one would expect an increase over time in therapy of the "Similarity of the Patient's Style of Expression to that of the Therapist" and of "Patient-Learning Discriminations," and no such increase (or decrease) in patient "Blocking" or "Negative Feeling Expression."

To evaluate this the ratings of the 40 samples for each class of patient behavior were grouped into five blocks across time-in-therapy (five raters for eight samples per block or 40 ratings per block) and the Grant Orthogonal Polynomial Trend Test Analysis of Variance (Grant, 1956) was used to test for the significance of components of trend. Further, *t* tests were used to test for significance of differences between early and late in therapy on all nine patient behavior classes. These data are presented in Table 2.

Of the classes of patient behavior to which the therapist selectively responded (i.e., reinforced), four out of five showed changes in patient behavior over time-in-therapy. Thus the data agree with the predictions in seven out of the nine classes of patient behaviors (78% correct prediction).

Considering the probability that the therapist also used other types of rewards or reinforcers and also rewarded other related patient behavior classes, considering the unknown differential complexity levels of the patient response classes, and considering the crudity of measurement, the findings strongly suggest that important reinforcement effects do indeed occur even in client-centered therapy.

Toward Evaluating the Validity of the Findings

There are, of course, some difficulties in interpreting the intercorrelation matrix. One might argue that these are simply interrelationships in the "heads" of the raters, as the raters might have known what the "X" value was when they rated a sample on "Y." However, each of the 12 variables was rated separately and they were rated in different orders. One would think it difficult to recall the X value of a given unit when the rating of the other units intervened between the X value and its corresponding Y value (an average of 240 ratings intervening between corresponding X and Y values). It could be argued that some of this bias is removed by the procedure for averaging the five different raters, since the raters were unaware of the actual hypothesis under study.

Beyond the above considerations, tabulation of one well-known characteristic of the therapist's behavior also suggests selective differential responding. The use of "uh huh" or "Mmm mmm" verbalizations has become, perhaps unfortunately, the hallmark of Rogerian psychotherapy. In the samples used in the present analysis, Mmm mmm's or Uh huh's occurred 23 times in a total of 12 of the 40 samples (in 30% of the samples). The Mmm mmm occurred in 9 of the 12 samples (75% of its occurrence) during high expression of negative feeling by the patient (all above the mean of ratings), while 0% occurred during low "patient negative feeling expression." In the remaining three samples, they occurred during the patient's direct restatement of what the therapist had just said. This tabulation alone suggests conscious or unconscious selective responding by the therapist, and is consistent with the obtained findings based upon relationships between rated therapist and patient classes of behavior.[4]

Finally, and most importantly, the obtained data dealing with changes in patient-in-therapy behavior were consistent with the obtained findings based upon prediction from a reinforcement view. Since the raters had no knowledge of whether a given sample came from early- or late-in-therapy, those findings of a tendency for significant linear increases to occur over time in reinforced patient behaviors and not to occur in nonreinforced patient behaviors, would also argue strongly against the notion that the obtained intercorrelations were simply "in the heads" of the raters.

IMPLICATIONS

The present findings point to the presence of significant differential reinforcement effects imbedded in the transactions of client-centered psychotherapy. Since differential reinforcement is one of the procedures used in operant research to alter (or control) behavior, the findings suggest that the therapist, in this case Rogers, implicitly alters (or controls) the patient's behavior in the therapeutic setting. To this extent, then, the evidence weighs in favor of the view proposed by Skinner rather than that of Rogers. The present findings are not consistent with Rogers' view that relatively *uniform conditions* which are globally "facilitative of personal growth and integration," are offered to patients in a manner not contingent upon the patient's behavior.

The present data, by demonstrating the role of empathy and warmth as positive reinforcers, suggest that the available evidence relating levels of these therapeutic conditions to patient outcome in therapy does not argue against a reinforcement interpretation of psychotherapy. On the contrary, the finding that empathy and warmth act as reinforcers suggests that the evidence relating empathy and warmth to patient outcome is open to a behavioristic interpretation, based in part on the therapist's use of differential reinforcement.

Recent studies have suggested that such humanistic qualities as empathy and warmth are antecedents to patient personality or behavioral change. In attempting to understand *how* such therapist qualities operate in producing therapeutic change, the present data suggest the potential value of studies utilizing behavioristic models. Since the available evidence relating empathy and warmth to patient outcome deals primarily with differences in *intensity levels* contaminated by differences in *rates* between therapists, it seems likely that additional and more precise understanding of the role of empathy (and hence more effective practice) might grow out of studies carried out from a reinforcement frame of reference. Considering only empathy as the type of reinforcer used in psychotherapy, it would be expected that successful and nonsuccessful therapists might differ in: (*a*) the particular patient behaviors chosen for differential reinforcement (say, self-concept statements versus historical-genetic statements); (*b*) the differential rate of reinforcement (say, 25% versus 75% for a specific class of patient behavior); (*c*) the intensity levels of the reinforcer used (say, the depth of empathy); and even the (*d*) scheduling of reinforcement (say, fixed ratio versus variable ratio).

---

[4] It should be noted that the therapist's use of the "Uh huh reinforcer" is relatively ineffective since there is no increase over time in "patient negative feeling expression."

Research aimed at identifying which patient behaviors, if reinforced at what intensity levels etc., lead to positive therapeutic outcomes would provide more specific knowledge of how such positive human qualities as empathy and warmth operate to produce personality or behavioral change in the patient.

Such an approach aims toward more specific knowledge, but not at all toward more mechanical therapy. As the communication of any "reinforcing machine" qualities would by definition mean a low level of empathy and warmth, the present viewpoint is in full agreement with Schonbar's (1964) statement that "as a therapist I am no more a 'reinforcing machine' than my patient is a 'talking pigeon.' "

REFERENCES

BANDURA, A. Psychotherapy as a learning process. *Psychological Bulletin,* 1961, **58,** 143–159.

BARRETT-LENNARD, G. T. Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs,* 1962, **76**(43, Whole No. 562).

BERGIN, A. E., & SOLOMON, SANDRA. Personality and performance correlates of empathic understanding in psychotherapy. Paper read at American Psychological Association, Philadelphia, September 1963.

CARTWRIGHT, ROSALIND D., & LERNER, BARBARA. Empathy: Need to change and improvement with psychotherapy. *Journal of Consulting Psychology,* 1963, 27, 138–144.

DICKENSON, W. A., & TRUAX, C. B. Group counseling with college underachievers: Comparisons with a control group and relationship to empathy, warmth, and genuineness. University of Kentucky and Kentucky Mental Health Institute, 1965.

DYMOND, ROSALIND. A scale for the measurement of empathic ability. *Journal of Consulting Psychology,* 1949, 13, 127–133.

EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika,* 1951, **16,** 407–424.

EYSENCK, H. J. The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology.* 1952, **16,** 319–324.

EYSENCK, H. J. The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1960. Pp. 697–725.

GRANT, DAVID A. Analysis of variance tests in the analysis and comparison of curves. *Psychological Bulletin,* 1956, **53,** 141–154.

HALKIDES, GALATIA. An investigation of therapeutic success as a function of four variables. Unpublished doctoral dissertation, University of Chicago, 1958.

HOBBS, N. Sources of gain in psychotherapy. *American Psychologist,* 1962, **17,** 741–747.

JOURARD, S. I-thou relationship versus manipulation in counseling and psychotherapy. *Journal of Individual Psychology,* 1959, **15,** 174–179.

KRASNER, L. The therapist as a social reinforcement machine. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy.* Vol. II. Washington, D. C.: American Psychological Association, 1962.

LESSER, W. M. The relationship between counseling progress and empathic understanding. *Journal of Counseling Psychology,* 1961, 8, 330–336.

ROGERS, C. R. *Client-centered therapy.* Cambridge, Mass.: Riverside Press, 1951. Pp. 73–74.

ROGERS, C. R. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology,* 1957, 21, 95–103.

ROGERS, C. R. *On becoming a person.* Cambridge, Mass.: Riverside Press, 1961.

ROGERS, C. R. The interpersonal relationship: The core of guidance. *Harvard Educational Review,* 1962, 32, 416–429.

ROGERS, C. R., KIESLER, D., GENDLIN, E. T., & TRUAX, C. B. *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics.* Madison: Univer. Wisconsin Press, 1965, in press.

ROGERS, C. R., & SKINNER, B. F. Some issues concerning the control of human behavior. *Science,* 1956, **124,** 1057–1066.

ROGERS, C. R., & TRUAX, C. B. The therapeutic conditions antecedent to change: A theoretical view. Chapter in, *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics.* Univer. Wisconsin Press, 1965, in press.

SCHONBAR, ROSALEA ANN. A practitioner's critique of psychotherapy research. Paper read at American Psychological Association, Los Angeles, September 1964.

STRUPP, H. H. Nature of psychotherapists' contribution to the treatment process. *Archives of General Psychiatry,* 1960, 3, 219–231.

TRUAX, C. B. Clinical implementation of therapeutic conditions. In Carl R. Rogers (Chm.), Therapeutic and research progress in a program of psychotherapy research with hospitalized schizophrenics. Symposium presented at the American

Psychological Association, New York, September 1961. (a)

TRUAX, C. B. The process of group psychotherapy. *Psychological Monographs,* 1961, **75**(7, Whole No. 511). (b)

TRUAX, C. B. Effective ingredients in psychotherapy: An approach to unraveling the patient-therapist interaction. *Journal of Counseling Psychology,* 1963, **10**, 256–263.

TRUAX, C. B., & CARKHUFF, R. R. For better or for worse: The process of psychotherapeutic personality change. Chapter in, *Recent advances in the study of behavioral change.* Montreal: McGill Univer. Press, 1963. Pp. 118–163.

TRUAX, C. B., & CARKHUFF, R. R. Significant developments in psychotherapy research. In Abt & Riess (Eds.), *Progress in clinical psychology.* New York: Grune & Stratton, 1964. Pp. 124–155.

TRUAX, C. B., CARKHUFF, R. R., & KODMAN, F., JR. Relationships between therapist-offered conditions and patient change in group psychotherapy. *Journal of Clinical Psychology,* in press.

TRUAX, C. B., WARGO, D. G., & SILBER, L. D. Effects of high conditions group psychotherapy with female juvenile delinquents. University of Kentucky and Kentucky Mental Health Institute, 1965.

WARGO, D. G. The Barron Ego Strength and LH [4] scales as predictors and indicators of change in psychotherapy. *Brief Research Reports,* 1962, **21**. (University of Wisconsin, Wisconsin Psychiatric Institute.)

WHITEHORN, J. C., & BETZ, BARBARA J. A study of psychotherapeutic relationships between physicians and schizophrenic patients. *American Journal of Psychiatry,* 1954, **3**, 321–331.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford Univer. Press, 1958.

(Received May 22, 1964)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 10–16

AUTONOMIC RESPONSIVITY TO VARIABLE INPUT RATES AMONG SCHIZOPHRENICS CLASSIFIED ON THE PROCESS–REACTIVE DIMENSION

WILLIAM D. WARD AND WILLIAM A. CARLSON

Veterans Administration Hospital, Danville, Illinois

Schizophrenic groups rated low (reactive), intermediate, and high (process) on the Elgin Prognostic Scale and a group of normal Ss discriminated between matched and unmatched pairs of nonsense syllables under varying rates of input. Continuous PGR measures were recorded as the Ss responded to the syllable pairs. Stimulus overload (defined as the extent to which the Ss were unable to respond appropriately) was assessed and shown to be consistent across groups. The groups were differentiated as to PGR responsivity. Among the schizophrenic Ss, the reactive group was the most responsive and the process Ss were least responsive to increases in rate of input.

A number of investigators have suggested that autonomic activity level is less for process than for reactive schizophrenic patients (identified by the Elgin Prognostic Scale; Becker, 1959; King, 1958; Pearl, 1962). Such studies have not taken into account the systematic manipulation of stimulus input at the time autonomic measures are taken. The developing theories relating autonomic responsivity, stimulus input, and personality development (Berlyne, 1960; Hunt, 1963; Pearl, 1962) encourage this latter refinement in such a line of investigation.

The theoretical formulations of Hunt (1963) and Pearl (1962) are particularly relevant to the relationships among schizophrenic pathology, stimulus input, and autonomic responsivity. The formulations of these authors suggest that process (Pr) schizophrenics should show less autonomic responsivity to stimulation than reactive (Re) schizophrenics because of improper toning of the reticular activating system during the developmental years. This deficiency in development is considered to be related to a prolonged state of stimulus deprivation. The notions of Hunt and Pearl also suggest that Pr schizophrenics would have a greater tendency to use perceptual defenses to shut out excessive input, that is, to reduce overload (OL). Within such a theoretical framework, observed differences in autonomic responsivity among Pr and Re schizophrenics might be explained as related to a response problem and/or to an input problem in Pr schizophrenics.

Gromoll (1961) used EEG indices in response to sudden, intense auditory and visual stimuli, and found no significant arousal response differences between Pr and Re schizophrenics. Perhaps because of the sudden, intense nature of the stimuli, Gromoll's results may reflect an inability on the part of his Pr subjects (*S*s) to institute perceptual defenses when overloaded. His study might suggest that there is no autonomic responsivity difference between Pr and Re schizophrenics if input is equal and not reduced by perceptual defense.

If Pr and Re *S*s were presented with material at variables rates, and the Pr *S*s showed lower autonomic responsivity, would such a response difference be the result of an autonomic response deficiency per se among the Pr *S*s or would it be the result of an input deficiency related to perceptual defenses against OL? The present study examined this problem by manipulating rate of input and by observing autonomic responsivity among Pr and Re schizophrenics under conditions which demonstrated that input was consistent across groups.

PGR was chosen as the autonomic index as it had previously differentiated between psychiatric diagnostic groups, and between neurotics under different conditions of stimulus input (Ward & Carlson, 1964). It was hypothesized that the Pr *S*s would demonstrate lower PGR responsivity to variable input rates than would the Re *Ss*.

METHOD

Subjects

Thirty-six male, hospitalized, schizophrenic veterans without neurological involvement were selected on the basis of Elgin Prognostic Scale scores, length of present hospitalization (LPH), age, and ward placement. Twelve normals were also included in the sample. The Elgin Prognostic Scale was originally developed by Wittman (1941); a revision by Becker (1959) was used in the present study. Clinical records and patient interviews were the sources of information on which the Elgin Scale ratings were based. Elgin interrater reliability, based upon 20 randomly selected cases, was shown to be .93. Two raters interviewed the 20 *Ss* jointly but examined patient histories and did their ratings independently. The interview material was not discussed by the raters until after the ratings were done.

The LPH was the number of months since a patient's last admission to any neuropsychiatric hospital. No effort was made to take into account the length of time the patients had been on trial visits and leaves of absence. Open ward patients ranging in age from 28 to 48 who had been hospitalized for at least a month were selected. The Elgin Scale distribution was arbitrarily divided into three groups: scores of 0 through 20 comprised the reactive group (Re), 21 through 30 the intermediate reactive-process group (RP), and the process *Ss* (Pr) were those from 31 through 50. The Re group included 5 paranoid, 5 undifferentiated, 1 catatonic, and 1 schizo-affective *S*. The RP *Ss* consisted of 6 paranoid, 5 undifferentiated, and 1 catatonic schizophrenic *S*. The Pr group contained 4 paranoid, 5 undifferentiated, 1 catatonic, 1 hebephrenic, and 1 simple schizophrenic *S*. Table 1 contains the means and standard deviations for the age and LPH variables. Group differences in these variables were assessed with simple analyses of variance. Although the Re *Ss* showed a trend toward shorter periods of hospitalization, as compared with the other two groups, neither age nor LPH group differences approached significance.

Selection criteria of a minmium of 1 month LPH and open ward placement were used on the assumptions that the groups would be similar in level of present functional adequacy and that medication effects would be comparable. The control for medication effects is difficult under the best circumstances. Little is known of the individual differences in reaction to drugs, "untranquilized" psychiatric patients are rather rare, and withdrawal of patient medication for research purposes might introduce an even greater source of error than ignoring the variable would introduce (Chapman, 1963). It was hoped that a patient on an open ward after a full month of treatment would have had his medication adjusted to his optimum maintenance dosage. Only one patient (Pr) was not on medication. The number of patients in each group who were on one drug were: Re, two; RP, seven and Pr, eight. The number on two drugs were: Re, nine; RP, four; and Pr, three. One Re *S* was on three drugs and one RP *S* was on four. Although the quantification of drug dosages has little meaning, Table 2 is presented for the reader's scrutiny. The normal *Ss* were nursing assistants selected so as to approximate the age distribution of the schizophrenic *Ss*. They were not on any type of medication and had had no psychiatric hospitalization.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ELGIN AND SUPPLEMENTARY VARIABLE MEASURES

Measure	Normals	Reactives	Reactive-Process	Process
Elgin		16.92 (2.47)	25.42 (1.73)	34.25 (2.80)
Age (yrs.)	38.17 (6.32)	36.42 (6.99)	41.00 (5.76)	36.92 (4.81)
LPH (mths.)		18.33 (25.72)	42.58 (39.73)	47.17 (41.08)
OL 2 seconds	2.33 (2.67)	4.00 (5.91)	7.33 (11.03)	9.00 (14.18)
OL 1 second	55.50 (22.74)	84.83 (23.13)	107.33 (41.40)	100.67 (29.12)
OL ½ second	121.67 (5.57)	120.50 (2.64)	120.25 (.97)	122.25 (6.57)

Note.—Standard deviations appear in parentheses.
LPH: length of present hospitalization; OL: measure of overload.

Task

The *Ss* were given series of nonsense syllables appearing on a two-window memory drum operating at a different speed for each experimental trial. Two syllables, one in each window, appeared simultaneously. The Ss were instructed to respond to matched pairs and were instructed not to respond to unmatched pairs. The syllables were randomly ordered in each list and half of the paired presentations were matched. The serial positions of the matched pairs were randomly determined. Three 2-minute adaptation trials and three 2-minute experimental trials were presented. The exposure time for each of the adaptation trials was 4 seconds per pair of syllables, and the exposure times for the experimental trials were, in order, 2 seconds, 1 second, and ½ second. During each of the adaptation trials, the same 30 pairs of syllables were presented once. During the experimental trials, a different list was used for each trial. For the 2 seconds, 1 second, and ½ second experimental trials, 30 pairs of syllables were presented twice, 4 times, and 8 times, respectively. Learning effects resulting from successive presentations of material were not crucial because of the method used in measuring OL, described below. One minute elapsed between trials.

TABLE 2

DRUG DISTRIBUTION AND DOSAGES

Drug	*S*s on each drug			Mean dosage (mg.) per day		
	Re	RP	Pr	Re	RP	Pr
Thorazine	8	6	7	562.50	533.33	685.71
Kemadrin	6	2	2	12.50	7.50	15.00
Stellazine	3	6	4	26.67	21.67	17.50
Tofranil	3	3		50.00	58.33	—
Mellaril	1	1		150.00	300.00	—
Librium	1			80.00	—	—
Benadryl		1		—	50.00	—

The *S*s were instructed to press a button, conveniently located on the right arm of their chair, each time a match appeared. They were instructed not to respond to unmatched pairs. The button activated a low intensity buzzer and activated a polygraph stylus which marked on the polygraph paper the precise points during the trials at which the *S*s responded. From an examination of the PGR tracings, it was obvious that neither the sound of the buzzer nor the *S*s' finger movements influenced the polygraph recordings.

Overload Measure

The degree of OL was defined as the extent to which a *S* was unable to respond appropriately to the stimuli. The OL score was a task efficiency measure which would have reflected group differences in learning effects and group differences in input related to perceptual defenses or attention problems. Groups that would have had attention problems or that would have used perceptual defenses would have tended to refrain from responding, would have tended to respond randomly, or would have tended to respond to every item. In all three cases, as explained below, the OL measure would approach a chance score according to the extent to which a *S* did respond intelligently.

A pair of syllables was considered to have been responded to if the point at which the *S* *initiated* action in depressing the button occurred while that pair of syllables had been exposed. Responding to an unmatched pair and not responding to a matched pair were considered to be incorrect. Since the duration of each trial was 2 minutes, the 2-seconds, 1-second, and ½-second trials involved the presentation of 60, 120, and 240 pairs of syllables, respectively. During the ½-second trial, 4 times as many items were presented as were presented during the 2-seconds trial and twice as many as were presented during the 1-second trial. The OL 2-seconds, OL 1-second, and the OL ½-second scores were equated by multiplying OL 2-Seconds Scores × 4 and OL 1-Second Scores × 2

A maximum OL score would approximate chance. Under the OL scoring system used, if a *S* responded to an item, the chance probability that he would be correct would be 1 in 2. If he did not respond, the same probability would hold. A chance score then would be 120 since the OL scores were based upon an assumed total of 240 presentations for each trial. The procedure of scoring OL as the sum of the number of unmatched syllables responded to and the number of matched syllables not responded to was useful in avoiding the possible confounding effect of individual differences in response tendency. If the OL score were merely the number of incorrect responses, a *S* who tended to respond excessively when overloaded might, by chance, get a lower OL score than a *S* who tended to respond less when equally overloaded.

Evidence supporting the method for scoring OL, and indicating that control of the response-tendency variable was accomplished, is suggested from an examination of the OL means under the ½-second trial. It was obvious that none of the *S*s could respond intelligently during this trial. According to the definition of OL, they were maximally overloaded. The OL means of the ½-second trial for all four groups approximated chance scores (120) even though there were significant differences between groups in the number of responses given. The fact that there were significant between group differences in number of responses suggested the possibility that the number-of-response variable could have confounded the Elgin-PGR responsivity relationship. Number-of-response by PGR responsivity correlations, however, eliminated this possibility; the correlations were on the order of zero. The PGR responsivity scores were difference scores derived by using the initial resistance score within each trial minus the lowest resistance measure following within that trial.

The measurement of OL would have reflected group differences in learning effects related to the successive presentations of the same lists of syllables within trials and would have reflected any group differences in input. The OL measure also eliminated response tendency as a possible confounding variable. It was assumed that if the four groups showed no differences in OL scores, learning effects and input would have been controlled. In comparing the OL means of the four groups under each trial, analyses of variance demonstrated that the 2-seconds and ½-second trial comparisons were well beyond the .20 level of significance. The 1-second trial analysis, however, showed a significant difference in OL among the four groups ($p < .005$). This significant difference in OL was obviously a function of the deviation of the normal group (see Table 1). A simple analysis of variance of the OL scores of the three schizophrenic groups (excluding the normals) did not approach significance ($p > .10$). The dependent variable analysis took the deviant 1-second trial OL scores of the normal group into account.

The three trials may be identified as involving three different degrees of OL. Using the entire sample, the results of a t test indicated that the OL 1-second mean was significantly different from the OL $\frac{1}{2}$-second mean ($p < .005$), and the OL 2-seconds mean was obviously different from the other two.

Procedure

The *S*s were brought to the Polygraph Laboratory individually. Twenty-five minutes elapsed from the time a *S* entered the room until the first experimental trial was begun. The first 10 minutes were spent in allowing the *S* to familiarize himself with the surroundings and in rapport-building conversation. The conversation was continued while the *S* was having the polygraph electrodes placed, after the experimental instructions were given. The last 10 minutes, before the experimental trials, were spent in three adaptation trials. In effect, the *S*s were told that the *E*s wanted to measure how aroused patients would become while taking a test that required quick responses. They were told that different kinds of patients would probably show different degrees of arousal, and this information might tell us how fast we could go in providing new therapy situations for different kinds of patients. Special emphasis was placed on statements that the procedure was not therapy and that any information relevant to therapeutic procedures, which might result from the study, would have implications for treatment only in the distant future. These statements were included in an attempt to minimize effects related to possible perceptions of the experimental situation as a therapeutic effort and effects related to possible anticipation of improved treatment.

All Ss were seen at either 2:30 or 3:30 P.M. in an air-conditioned room which maintained a suitably constant temperature of about 74 degrees. Corn plasters, placed on the second and third fingers of the left hand, held zinc electrodes bound by adhesive tape. The electrode paste was manufactured at the hospital pharmacy according to the specifications suggested by Lykken (1959). Continuous PGR records were taken with the use of a Grass, six-channel polygraph. After the three adaptation trials, the Ss were told that the exposure time would decrease with each successive trial. Between experimental trials, the lists of nonsense syllables were changed, and the *S*s were encouraged to relax and refrain from talking or moving excessively.

RESULTS

It was desired that the individual trials should be assessed independently, as well as dependently, so that it would be possible to identify differences in trends between groups under specific degrees of OL (none, moderate, or extreme). Separate analyses were done within trials and across trials.

The dependent-variable measures for the within-trials analyses were taken from the PGR scores for each of the three, 2-minute experimental trials. The PGR levels were sampled at time zero and every 30 seconds thereafter for each trial. Each *S* had five scores for each of the three trials, and the raw scores were resistance measures in millivolts. The summaries of the groups by points within-trials analyses of variance appear in Table 3, and the group trends may be seen in Figure 1. Figure 1 includes the third adaptation trial trends (4-seconds trial) for comparative purposes.

Tests for heterogeneity of variance were done for each trial. In all cases the hypothesis of heterogeneity was rejected ($p > .99$). The tests of main interest for the within-trials analyses were the Groups × Points within-

TABLE 3

SUMMARY OF GROUPS BY POINTS WITHIN TRIALS ANALYSES OF VARIANCE

Source	*df*	2″ Trial		1″ Trial		1/2″ Trial	
		MS	*F*	*MS*	*F*	*MS*	*F*
Groups (A)	3	556.35	—	312.77	—	316.77	—
Error (a)	44	762.70		607.20		595.82	
Points within Trials (B)	4	5.97	3.55*	25.54	16.06***	13.07	46.68***
A × B	12	4.52	2.69**	2.83	1.78	1.44	5.14***
Error (b)	176	1.68		1.59		.28	

* $p < .01$.
** $p < .005$.
*** $p < .001$.

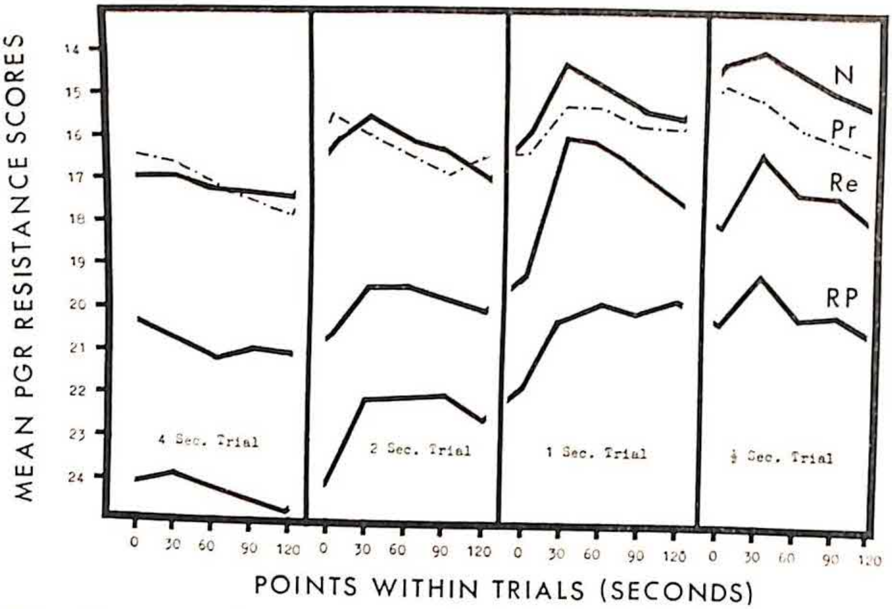


Fig. 1. The mean PGR resistance scores for each group taken every thirty seconds during each trial. (The lower the score, the higher the arousal.)

trials interactions which reflect group differences in arousal trends over each of the 2-minute trial periods. The greatest between group variance in trend for all experimental trials occurred during the first 30 seconds. The recovery trends, after the first 30 seconds in each of the three trials, follow a similar pattern in the four groups.

The Groups × Points within-trials interactions for the 2-seconds trial and the ½-second trial reached significance. The Pr *S*s were the only group that did not show a positive response during the first 30 seconds. During the 1-second trial, the trends of the four groups did not differ significantly ($p < .06$), and all groups showed a positive response during the first 30 seconds. A possible explanation for the failure of the 1-second trial Groups × Points within-trial interaction to reach significance is offered below in the Discussion section. As mentioned previously, the 2-seconds, 1-second, and ½-second trials represented no OL, moderate OL, and extreme OL, respectively.

As previously indicated, the normals were significantly less overloaded on the 1-second trial than were the other three groups, and the three schizophrenic groups did not differ significantly as to OL on the 1-second trial. An analysis of variance which excluded the normal *S*s, yielded a PGR, 1-second trial, Groups × Points within-trial interaction which did not approach significance ($p > .10$). It may be considered that the lower OL level of the normal *S*s on the 1-second trial did not confound the dependent variable analysis.

The relative magnitude of group responsivity during the first 30 seconds of each experimental trial maintained quite a consistent pattern across trials. For the 1-second and ½-second trials, group responsivity in descending order of magnitude was Re, RP, normal, and Pr. For the 2-seconds trial, the pattern was the same except that the Re and RP groups were reversed. Difference scores were computed for each *S* for each trial using his initial score minus his 30-seconds score. A multiple-range test for each of the three experimental trials indicated that the RP-Pr means for the 2-seconds trial differed at less than the .05 level, that no two of the four 1-second trial means differed significantly, and that on the ½-second trial, the Re-Pr means differed ($p < .01$) as did the RP-Pr means ($p < .05$).

The Pr *S*s showed a higher overall level of arousal than the other two schizophrenic groups. This raised the question as to whether the difference could reflect a chronic state of high arousal among the Pr *S*s or whether it could represent a reaction to the experimental situation as a whole. The apparent difference, however, was not significant. The within-groups variance was large enough that the between *S*s main effects for the Groups × Points within-trials analyses (see Table 3) did not approach significance. To further check this finding, simple analyses of variance were done for each of the four trials using the first score (0-second score) for each *S* in each trial. None of the four analyses approached significance. The overall arousal levels of the groups did not differ significantly.

It has been mentioned previously that there were no confounding group differences in OL on any of the trials, but as a check that the *S*s were in fact responding to changes in rate and not to the degree of OL, correlations were done between OL scores and PGR difference scores (initial point during a trial minus the 30-seconds point during that trial) for the 1-second and ½-second trials. The correlations were on the order of zero, indicating that there was no relation between number of items missed (OL) and PGR response. Scatter plots revealed random distributions.

The dependent variable measures for the Groups × Trials analysis of variance were derived by subtracting the lowset PGR resistance measure of the 30-seconds, 60-seconds, 90-seconds, and 120-seconds points within each trial from the 0-second score of the 2-seconds trial. Each *S* had one score per trial. The summary of the Groups × Trials analysis appears in Table 4. The test of main interest in this analysis was the Groups × Trials interaction. The findings tended to confirm the within trials analyses, that differences in arousal among groups arose from the differences in rates at which the pairs were presented in the different trials.

TABLE 4

SUMMARY OF GROUPS BY TRIALS ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*
Groups (A)	3	107.54	2.60
Error (a)	44	41.42	
Trials (B)	2	53.82	14.75**
A × B	6	8.51	2.33*
Error (b)	88	3.65	

* $p < .05$.
** $p < .001$.

DISCUSSION

The arousal responsivity of the normal *S*s presents an interpretive problem. As can be seen from Figure 1, the normals were more similar to the Pr *S*s in pattern within trials, in pattern across trials, and in gross level than they were to the Re or the RP group. An attempt to explain this phenomenon is beyond the scope of the present paper. One might speculate that the Pr and normal groups were low in responsivity for different reasons, but hypotheses regarding qualitative differences between the groups in task perception or task approach would be premature. The most reasonable approach to the findings of the present study is to confine all generalizations to the interrelationships among the schizophrenic groups.

The results of the present study indicated that groups of schizophrenic *S*s differing in Elgin Prognostic Scale scores can be differentiated as to PGR responsivity to variable rates of stimulus input. The effects were interpreted as reflecting responses to positive changes of rate of input rather than to the stress of overload. The same pattern of autonomic response was found as was found in previously mentioned studies which did not use variable rates of stimulus input as a manipulation. Within each experimental trial and across trials, the Pr *S*s were the least responsive with the normal group next to the lowest. In two out of three trials, the Re *S*s were the most responsive and the RP group was the next highest. The differentiation of the groups was in accord with the hypothesis; under variable rates of input, Pr schizophrenia was consistently associated with low autonomic responsivity. Support for the theoretical formulations regarding the relationship between developmental experiences with stimulus input and autonomic responsivity was suggested.

A speculative comment is in order regarding the within-experimental trials responsivity of the Pr *S*s. The normal, Re, and RP groups responded initially with increased arousal in each of the trials while the Pr *S*s showed adaptation virtually throughout the 2-seconds and the $\frac{1}{2}$-second trials and demonstrated increased arousal only on the 1-second trial. The positive response of the Pr *Ss* on the 1-second trial was probably responsible for the failure of the 1-second trial Groups × Points within-trial interaction to reach significance ($p < .06$). Perhaps this suggests something about optimum rate of input for Pr schizophrenics. If one may assume that some positive response is desirable, perhaps the optimum rate of input for Pr *Ss* was presented in the 1-second trial.

The relationship between the process-reactive dimension and PGR responsivity also supports the notions of Funkenstein, Greenblatt, and Solomon (1952) and of Pearl (1962), that there is a positive relationship between autonomic responsivity and prognosis among schizophrenics.

Further research along the lines of the present study is indicated. In extending the developmental implications of information processing, it would be well to relate the Pr-Re theory to other stimulus variables such as intensity and perceptual incongruity. The consideration of these variables in possible interactive relationships with rate of input is indicated. Therapeutic implications might also be pursued. Fruitful approaches in the therapeutic area might be concerned with the identification of optimum levels of autonomic response for behavioral efficiency and the identification of optimum rates of input.

REFERENCES

BECKER, W. C. The process-reactive distinction: A key to the problem of schizophrenia? *Journal of Nervous and Mental Disease,* 1959, 129, 442–449.

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

CHAPMAN, L. J. The problem of selecting drug-free schizophrenics for research. *Journal of Consulting Psychology,* 1963, 27, 540–542.

FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. An autonomic nervous system test of prognostic significance in relation to electroshock treatment. *Psychosomatic Medicine,* 1952, **14**, 347–362.

GROMOLL, H. F. The process-reactive dimension of schizophrenia in relation to cortical activation and arousal. Unpublished doctoral dissertation, University of Illinois, 1961.

HUNT, J. McV. Motivation inherent in information processing and action. In, O. J. Harvey (Ed.), *Motivation and social organization: The cognitive aspect.* New York: Ronald Press, 1963. Pp. 35–94.

KING, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *Journal of Abnormal and Social Psychology,* 1958, **56**, 160–164.

LYKKEN, D. T. Properties of electrodes used in electrodermal measurement. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 629–634.

PEARL, D. Stimulus input and overload in relation to classifications of schizophrenia. *Newsletter for Research in Psychology,* 1962, 4, 44–56. (Kecoughtan, Va.: VA Center)

WARD, W. D., & CARLSON, W. A. Cognitive dissonance, opinion change, and physiological arousal. *Journal of General Psychology,* 1964, **71**, 115–124.

WITTMAN, PHYLLIS. A scale for measuring prognosis in schizophrenic patients. *Elgin State Hospital Paper,* 1941, 4, 20–33.

(Received December 4, 1964)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 17–24

REFLECTION–IMPULSIVITY:

THE GENERALITY AND DYNAMICS OF CONCEPTUAL TEMPO[1]

JEROME KAGAN

Harvard University

3rd-grade children previously classified as either reflective or impulsive were administered a serial learning task under 3 different conditions. After administration of 2 lists, 1 group was told their performance was poor, a 2nd was told that the next lists were difficult, and controls were given no special communication. Impulsive Ss in all groups reported more *incorrect* words before and after experimental intervention. Reflective boys who were told the next lists were difficult showed the largest increase in incorrect words.

The dramatic differences in quality of problem solving among children of the same age or among children of different ages have been attributed primarily to two categories of constructs—motivational variables and/or adequacy of conceptual skills. In effect, differences in quality of cognitive products have been explained by assuming either that one child cared more about his performance, or that one child had more knowledge relevant to the task. However, other cognitive processes are an intimate part of problem-solving activity. One of these processes concerns the degree to which the child reflects over the adequacy of a solution hypothesis.

Some children—and adults—select and report solution hypotheses quickly with minimal consideration for their probable accuracy. Other children, of equal intelligence, take more time to decide about the validity of solutions. The former group has been called impulsive, the latter reflective. The reflection-impulsivity dimension exerts its influence at two places in the problem-solving sequence. A schematic description of the chronology of problem solving includes four phases plus a reporting phase.

Phase 1: Decoding of the problem; comprehension of the problem.

Phase 2: Selection of a likely hypothesis on which to act in order to arrive at solution.

Phase 3: Implementation of the hypothesis (e.g., carry out a relevant arithmetic operation; proliferate a series of synonyms or associates).

Phase 4: Evaluate the validity of the solution arrived at in Phase 3.

Phase 5: Report of solution to an external agent.

The reflection-impulsivity dimension is influential in Phases 2 and 4, the times when the child is selecting a hypothesis to work on mentally or when he is about to report an answer to a teacher, peer, or parent. Previous work has suggested that impulsive selection or reporting of hypotheses is associated with inaccurate performance when adequacy of the child's knowledge repertoire is controlled. Moreover, the tendency to be reflective or impulsive shows intraindividual stability over time and generality across situations involving visual recognition or matching problems (Kagan, in press; Kagan, Rosman, Day, Albert, & Phillips, 1964).

The major purpose of this paper is to explore the relation between the disposition to be reflective or impulsive, and accuracy of performance in a different kind of cognitive task. Specifically, this investigation examines the relation between impulsivity and errors of commission in a series learning task (i.e., reporting words that were never present on the original list). A preliminary study of third- and fourth-grade children revealed a positive relation between impulsivity and errors of commission and this study was a more elegant attempt at replication (Kagan, 1965).

A second purpose of this work is to inquire

---

[1] This research was supported in part by research grant MH-4464 and MH-8792 from the National Institute of Mental Health, United States Public Health Service. Part of this work was conducted while the author was at the Fels Research Institute. The author is indebted to Deborah Day, William Phillips, and Barbara Lobell for their assistance in this study.

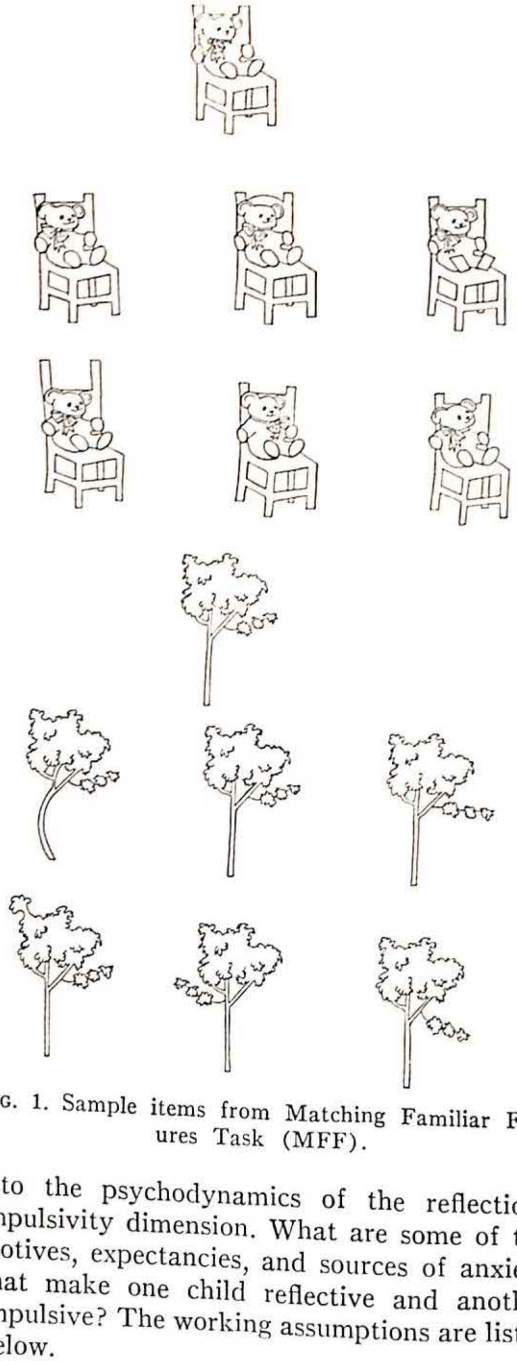


FIG. 1. Sample items from Matching Familiar Figures Task (MFF).

into the psychodynamics of the reflection-impulsivity dimension. What are some of the motives, expectancies, and sources of anxiety that make one child reflective and another impulsive? The working assumptions are listed below.

Assumption 1. The relation of reflection-impulsivity to quality of performance only obtains for problems that have response uncertainty (i.e., several response alternatives occur to the child simultaneously, or in close temporal contiguity, and the child must select the one he judges to be most valid).

Assumption 2. For problems that are difficult and have response uncertainty, the child's tendency to be reflective (i.e., have long decision times) or impulsive (short decision times) is a function of the balance between the strengths of the following two standards: produce the answer quickly versus do not make a mistake. The relation between the positive value of quick success and the anxiety generated by the possibility of committing an error determines the child's decision time (i.e., his tendency to be reflective or impulsive).

Assumption 3. If the child's anxiety over a possible error is much stronger than his desire for quick success, he will be reflective. If his anxiety over committing an error is weak in relation to his desire for quick success, he will be impulsive.

On the basis of these assumptions it was expected that situations that elicited anticipation of failure would lead to greater anxiety and greater task disruption in reflective than in impulsive children. Since psychology does not possess faithful indexes of these two constructs (i.e., desire for success, anxiety over failure), one possible strategy is to devise experimental interventions that are expected to elicit these states differentially and examine the differential effect on children previously classified as impulsive or reflective. The reader should note that *anxiety over possible failure* is psychologically distinct from *expectation of failure*. A child who has been exposed to chronic failure may enter into a problem situation with a strong anticipation of failure but minimal anxiety. Many adolescents who comprise the high school "drop out" group fit this description. They do not expect to perform with competence yet they appear to be accepting of this state of affairs. The literature on "fear of failure" has often failed to distinguish between expectancy of failure and anxiety over anticipated failure.

The Measurement of Reflection-Impulsivity

Previous research on reflection-impulsivity has used the child's performance on a visual recognition task (called MFF for Matching Familiar Figures) as the primary index of the child's position on this dimension. In this test the child is shown a single picture of a familiar object (the standard) and six similar variants, only one of which is identical to the standard. The child is asked to select the one variant that is identical to the standard. Figure 1 illustrates two sample items from the MFF test.

The critical variables scored are response time to the child's first answer and the total number of errors across the 12-item test. Impulsive children in Grades 1–4 have a mean response time between 4 and 10 seconds and make about 15–20 errors. Reflective children have mean response times between 30 and 40 seconds and make between 2 and 6 errors.

In sum, this study tested two hypotheses: (*a*) a positive association between impulsivity on the MFF test and errors of commission on a serial learning task, and (*b*) greater deterioration in serial learning performance for reflective, in contrast to impulsive, children to a communication that suggested the strong possibility of future failure.

METHOD

Subjects and Procedure

Third-grade subjects (*S*s) (136 boys and 107 girls) from two public schools in neighboring Ohio communities were seen individually for a 1-hour session during which the Matching Familiar Figures and WISC vocabulary and information scales were administered. The *S*s were tested by either a male or female examiner.

Selection of groups. From this pool of 243 *S*s, smaller groups of reflective or impulsive children were selected. The basis of categorization of an *S* as reflective or impulsive included both response time and errors on the Matching Familiar Figures task. The *S*s classified as reflective were above the median on MFF response time and below the median on MFF errors for their sex. Impulsive *S*s were below the median on MFF response time and above the median on MFF errors. The reflective and impulsive *S*s were then assigned to one of three groups—a threat group, a rejection group, and a control group. The reflective and impulsive children assigned to these groups were matched on sex and scaled score on the WISC subtests. Thus the mean and range of verbal ability were equal for the reflective and impulsive *S*s across all three experimental groups. Assignment of *S*s in this 2 × 2 × 3 design (12 cells) yielded sample sizes varying from 13 to 24 *S*s in the 12 cells. There were 53 reflective boys and 65 impulsive boys; 45 reflective girls and 40 impulsive girls.

Serial learning procedure. Each *S* was seen 3 to 4 months after the initial session by a male experimenter (*E*) for administration of the serial learning task. The detailed procedure follows:

The *E* told the child he was going to test his memory and the child was urged to do as well as he could. The exact instructions were:

TABLE 1

LISTS OF WORDS USED IN SERIAL LEARNING TASK

List 1	List 2	List 3	List 4
THUNDER	SHOE	PLATE	CHAIR
DOG	WATER	RIVER	WHEEL
LIGHTNING	COAT	BREAD	COUCH
PENCIL	ROPE	PIPE	BALL
RAIN	SHIRT	SPOON	TABLE
BARN	FLOWER	TREE	FINGER
KNIFE	PANTS	STOVE	BED
STORM	NAIL	CAR	GLUE
LEAF	DRESS	CUP	DESK
WIND	PAPER	DOOR	FIRE
HAMMER	SWEATER	NAPKIN	LAMP
	IRON	SNOW	SOAP

> You will hear someone read some words to you on this tape recorder. Try to remember all the words you hear for I will ask you to say the words back to me. The words will be read slowly. Let's listen to a practice list of words.

The *E* allowed the *S* practice with two lists of three and four words each until the *S* understood the nature of the task. The criterion task consisted of four different lists of 12 familiar words each. The lists were recorded at a rate of 1 word every 4 seconds using the author's voice. Each *S* was given two trials on each list. The second trial for each list presented the words in an order that was the exact reverse of the first trial. A recall was obtained after each trial, providing two recall scores for each list. Table 1 contains the list of words.

Each of the lists contained six words that belonged to a conceptual category, but each of these conceptually related words was surrounded by a word unrelated to that concept. For example, the List 1 concept included words associated with a "rainstorm" (i.e., thunder, lightning, rain, puddle, storm wind). The List 2 concept was concerned with "articles of clothing" (i.e., shoe, coat, shirt, pants, dress, and sweater). The List 3 concept included words connected with "eating" (i.e., plate, bread, spoon, stove, cup, napkin). The List 4 concept was "furniture" (i.e., chair, couch, table, bed, desk, lamp). The six remaining words in each list were minimally related to each other or to the concept contained in that list.

All *S*s were treated alike for the first two lists. The three different experimental treatments occurred following List 2.

The threat group was told:

> O.K., that was fine. These two lists we just finished were really practice lists to give you an idea of what this kind of test is like. The next two lists really count. You will have to do very well on these next two lists. They are hard and children who don't have good memories usually don't do very well, so try and concentrate.

The intent of this communication was to arouse anxiety over possible failure.

The children in the rejection group were told:

> That was poor, you are not remembering enough words. You should do better than that. You should have remembered more words than you did.

The intent of this communication was to arouse anxiety over the examiner's disapproval of the child's performance.

The children in the control group were told:

> O.K., let's take a short rest and then do some more lists.

It is acknowledged that all children may not have reacted to the instructions in the manner hypothesized, for individual differences in response to a communication are always to be expected. However, it is assumed that the majority of *S*s in each group experienced thoughts and feelings appropriate to the communication, and that more *S*s in the threat and rejection groups experienced anxiety than *S*s in the control group.

Following the experimental communication each *S* was then administered Lists 3 and 4. The major variables scored for both trials of each list were: (*a*) total number of concept words recalled, (*b*) total number of nonconcept words recalled, (*c*) number of intrusion errors, when an intrusion error was defined as the reporting of a word that was not present on the original list read to the child.

RESULTS

Intercorrelation of Recall and Intrusion Scores

Complete intercorrelation matrices were computed for each of the recall and intrusion error scores for the sexes separately and for reflective and impulsive *S*s separately for the first two lists, before the experimental interventions.

The relation between recall of concept and nonconcept words (pooling the first two lists) was high for reflective *S*s but low for impulsives ($r = .36$ and .55 for reflective boys and girls, respectively; $r = .19$ and .18 for impulsive boys and girls).[2] Reflective *S*s were more consistent in their recall scores for concept and nonconcept words. There was a moderate relation between Lists 1 and 2 for the recall of concept words ($r = .40$, .24 for reflective and impulsive boys; $r = .53$ and .38 for reflective and impulsive girls). For nonconcept words the corresponding coefficients were .30 and .28 for boys; .20 and .27 for girls. There was a low, positive relation between Lists 1 and 2 for intrusion errors ($r = .13$, .37 for boys; .24 and .10 for girls). These correlations were low because the means and standard deviations for intrusion errors were relatively low.

The relationship between recall and intrusion errors was generally negative—the higher the recall score the lower the number of intrusions. The correlations for List 1 were −.40 and −.30 for reflective and impulsive boys; .02 and .−.37 for reflective and impulsive girls. For List 2, the corresponding correlation coefficients were .08 and −.40 for boys and −.25 and .07 for girls. Thus, the most consistent negative relationships between recall and intrusions for the first two lists held for impulsive boys.

Relation to WISC Scores

The relationship of WISC verbal skills to recall and intrusion scores is of interest. There was no relationship between verbal ability and intrusion errors for either sex or for reflective or impulsive *S*s. The coefficients ranged from −.16 to +.12, and none of the coefficients was significant. Thus, as with response time on MFF, an impulsive attitude (in this case measured by intrusion errors) was orthogonal to verbal resources.

Verbal skills were related to recall in a complicated manner. Among boys, there was no relationship to recall of concept related words ($r = .10$ and .08); but there was a moderately positive relation with recall of nonconcept words ($r = .41$ for reflective boys and .20 for impulsive boys). Verbal skills facilitated the recall of words having no associative link among them, but had no such facilitating effect on words that were conceptually related to each other. Although this result may appear puzzling, it is not so enigmatic if one views the measure of verbal ability as reflecting, in part, motivation to master intellectual tasks. It is more difficult to recall unrelated words than related words and the greater the motivation to perform well on the learning task the higher this score should be. This increased motivation should also be reflected in higher recall scores for concept words. But the recall score for concept words was considerably higher than that of nonconcept words (a mean of 15 versus 11 for Lists 1 and 2), and it may be that all the children were performing at close to their intellectual capacity for the conceptually related words. As a result, added motivation was less facilitating for these words. It is possible that the brighter boys recalled more nonconcept words because their richer verbal repertoire allowed them to create connections between the unrelated words. However, it is difficult to imagine how a third-grader might quickly connect words like dog-pencil-barn-knife-leaf-hammer as they were being read to him at a relatively fast rate.

---

[2] The value of r for significance at .05 or better (two tailed) is .27 for reflective boys, .24 for impulsive boys, .29 for reflective girls, and .30 for impulsive girls.

Among girls there was no dramatic difference in the association between verbal skills and recall of concept or nonconcept words. The correlations with concept word recall were .59 and .24 for reflective and impulsive girls; the corresponding correlations for nonconcept words were .48 and .47. Thus brighter girls recalled more of both classes of words. This finding is also reflected in the dramatic sex difference in degree of association between verbal skills and total recall for all words for the first two lists. The correlations were high for girls ($r = .61$ and .46 for reflectives and impulsives), but lower for boys ($r = .31$ and .18). This type of result is not unusual in our work. It is the rule rather than the exception that verbal ability is typically more highly correlated with a learning or performance measure for girls than it is for boys. Knowledge of a girl's IQ allows one to predict her performance on an intellectual task with greater accuracy than is possible from knowledge of a boy's IQ score. In sum, recall and intrusion scores displayed positive but low generality across the first two lists and recall scores were negatively related to intrusion errors.

The Superiority of Concept Recall

The superior recall of concept over nonconcept words for all *Ss* for the first pair of lists supports the reasonable assumption that the availability of a mediational link for a group of words facilitates recall of members of the mediationally related class. The differential recall of concept over nonconcept words was assesed for each *S* for the first and second pairs of lists by subtracting the latter score from the former for each *S*. The differences for Lists 1 and 2 were dramatic, for 75% of the boys and 80% of the girls recalled more concept than nonconcept words ($p < .001$ for each sex). The differences between concept and nonconcept recall were not as dramatic for Lists 3 and 4.

Relationship of Recall and Intrusions to Reflection-Impulsivity

The mean scores for recall and intrusions for Lists 1 and 2 are presented in Table 2 for each of the 12 groups. The maximal recall score for concept or nonconcept lists for either

TABLE 2

MEAN SCORES ON SERIAL LEARNING VARIABLES

Variable	Boys						Girls					
	Reflective			Impulsive			Reflective			Impulsive		
	Threat	Reject	Cont.	Threat	Reject	Cont.	Threat	Reject	Cont.	Threat	Reject	Cont.
Concept Recall Lists 1 & 2	15.1	15.8	15.6	14.0	13.6	14.0	15.4	14.8	14.7	13.9	14.4	14.6
Nonconcept Recall Lists 1 & 2	12.8	10.8	12.9	11.0	10.9	11.3	12.2	12.2	11.4	11.2	11.4	11.4
Concept Recall Lists 3 & 4	11.5	12.6	12.6	12.1	13.5	11.1	13.4	13.7	11.2	13.1	13.3	12.3
Nonconcept Recall Lists 3 & 4	11.5	12.2	12.3	10.1	11.8	10.9	12.1	11.9	11.6	10.5	12.4	10.2
Intrusions Lists 1 & 2	1.2	1.1	1.4	1.7	2.0	2.9	1.0	1.2	1.5	1.5	2.1	2.0
Intrusions Lists 3 & 4	5.8	3.2	3.1	4.3	3.9	5.2	3.2	2.3	1.6	3.6	3.3	3.1

8-423
9.70

the first two lists or the second two lists was 24 words.

Analyses of variance were performed for each of the six major variables. In order to expedite these analyses *S*s were randomly eliminated from each group containing more than 13 *S*s in order to have a constant sample size of 13 for each of the 12 cells. The means and standard deviations for each of the 12 cells were proportional to the parameters derived from the slightly larger samples whose means were presented in Table 2. Summaries of these analyses of variance appear in Table 3.

Recall data. Reflective *S*s recalled more words than the impulsives, especially for Lists 1 and 2. The superior recall of reflective *S*s on the first two lists was significant for the nonconcept words ($t = 2.98$, $p < .01$), and also significant for the nonconcept words ($t = 2.22$, $p < .05$). The mean recall scores for Lists 3 and 4 were lower for all groups, including the controls. Thus the poorer recall scores on the second pair of lists cannot

TABLE 3

ANALYSES OF VARIANCE

Source	*df*	*SS*	*MS*	*F*	*p*
Concept words recalled: Lists 1 & 2					
Variable					
Reflection (R-I)	1	73.4	73	7.52	<.01
Sex (S)	1	0.2	0.2		
Error	153	1,485	9.70		
Nonconcept words recalled: Lists 1 & 2					
R-I	1	37.3	37.3	4.78	<.05
S	1	1.2	1.2		
Error	153	1,194	7.8		
Concept words recalled: Lists 3 & 4					
R-I	1	18	18		
S	1	15.4	15.4		
Conditions (C)	2	61.2	30.6	3.03	<.05
R × S	1	5.4	5.4		
R × C	2	6.2	3.1		
C × S	2	17.9	8.9		
R × C × S	2	15.6	7.8		
Error	144	1,459	10.1		
Nonconcept words recalled: Lists 3 & 4					
R	1	45.2	45.2	5.48	<.05
S	1	0.2	0.2		
C	2	57.5	28.7	3.49	<.05
R × S	1	2.1	2.1		
R × C	2	16.7	8.3		
C × S	2	2.6	1.3		
R × S × C	2	0.8	0.4		
Error	144	1,187	8.25		
Intrusion errors: Lists 1 & 2					
R	1	27.1	27.1	7.38	<.01
S	1	2.8	2.8		
Error	153	576	3.7		
Intrusion errors: Lists 3 & 4					
R	1	79.0	79.0	4.97	<.05
S	1	100.1	100.1	6.25	<.05
C	2	9.6	4.8		
R × S	1	8.8	8.8		
R × C	2	80.4	40.2		
C × S	2	18.7	9.4		
R × S × C	2	27.5	14.7		
Error	144	2,290	15.9		

be attributed to the different experimental communications.

The superior recall of reflective *S*s does not appear to be a function of either longer delays before beginning to report words or the character of the initial words reported. There was no significant difference between reflective and impulsive *S*s on response latency to begin to report words after termination of the taped list. Moreover, detailed analysis of the first five words recalled by each *S* did not reveal a greater tendency for reporting concept words by reflective children. The initial five words reported tended to obey the laws of primacy and recency. Thus the superior recall of reflective children seems due to the fact that they persisted longer in their attempt to produce a better cognitive product, suggesting that they were more highly motivated on this task. This suggestion matches other data on reflective and impulsive children which indicates that reflective *S*s persist longer with difficult intellectual tasks than their impulsive counterparts with the same verbal ability (Kagan, 1965).

The effect of conditions on recall for Lists 3 and 4 suggested a tendency for the control *S*s to have lower recall scores than either the threat or rejection groups. Examination of Table 3 reveals that the "rejection" statement acted as an incentive for most of the *S*s.

Intrusion errors. A major purpose of this study was to replicate an earlier study in which impulsive children produced more intrusion errors than reflectives. These data clearly verified this association for all four lists. The analysis of variance for Lists 1 and 2 revealed a significant F ratio ($F = 7.38$, $p < .01$) for the reflection-impulsivity dimension, with impulsive *S*s producing more intrusion errors than their reflective counterparts. On Lists 3 and 4, the effect of the reflection-impulsivity dimension was still present ($F = 4.97$, $p < .05$), and a sex difference occurred, with boys producing more intrusion errors than girls ($F = 6.25$, $p < .05$). This sex difference is also revealed in the results of an additional analysis. Specifically, the change in intrusion errors from List 2 to List 3 was computed for each S and a constant of 10 added to avoid negative scores. The analysis yielded only one significant F ratio ($p < .01$), with boys displaying larger increases in intrusion errors than girls.

The effect of threat on reflective Ss. There was the suggestion of the predicted association between reflection and the threatening communication when the increase in intrusions from List 2 to List 3 was the score analyzed ($F = 2.56$, $p < .10$ for the interaction between reflection-impulsivity and experimental group). Specifically, reflective children following threat showed the largest increase in intrusion errors, while reflective *S*s in the control group produced the smallest increase in intrusions. Moreover, the data in Table 2 reveal that the largest increase in intrusion errors, comparing Lists 3 and 4 with Lists 1 and 2, occurred among reflective boys under threat conditions. These boys showed an increase of 4.6 intrusions, compared with values of 2.1 and 1.7 for reflective boys under rejection and control conditions. A test of the significance of the difference between the threat and rejection groups (4.6 versus 2.1) yielded a t of 1.70 ($p < .10$, two tails). Furthermore, the increase in intrusion errors comparing Lists 3 and 4 with Lists 1 and 2 was greater for the reflective boys under threat than the impulsive boys under threat (4.6 versus 2.6), but this difference was not statistically significant. The frequency data indicated that 60% of the reflective boys under threat showed an increase of 3 or more intrusion errors and no reflective boy displayed a decrease. Among the impulsive boys, only 38% had an increase of 3 or more intrusions and 3 boys had fewer intrusions after threat than they did before threat. Although these differences are not statistically significant by conventional standards, they support the notion that reflective boys were more influenced by the instruction that suggested the possibility of future failure.

The differential deterioration in recall also supports these ideas. When concept and nonconcept words were pooled, the reflective boys under threat recalled 4.9 fewer words on Lists 3 and 4 than they did on Lists 1 and 2. This difference compares with a value of 2.8 for impulsive boys under threat. The recall loss of 4.9 words between the first and second pairs of lists is the largest average drop in recall displayed by any of the 12 groups (average loss of 2.0 words for remaining 11 groups). Assuming that a combination of a large increase in intrusion errors and a large decrease in recall would be most indicative of deterioration and, by inference, of anxiety, one final analysis was attempted. Subjects who showed both an increase in intrusion errors of three or more (the median value of all *S*s) and a decrease in recall of four or more words were selected. Forty percent of the reflective boys under threat fulfilled this double index of deterioration in performance, in contrast to 27% among impulsive boys. Moreover, 4 of the 15 reflective boys under threat deteriorated 12 or more words comparing Lists 3 and 4 with recall in the first two lists. The largest loss in recall among the impulsives was 9 words. Although these differences missed statistical significance, they agree with the notion that telling the child that a task he was about to do was difficult—and failure a likely possibility—had a greater disruptive effect on reflective than impulsive boys.

DISCUSSION AND CONCLUSION

The hypothesis that children classified as impulsive on the visual matching test would make more errors of commisison in the serial learning procedure was clearly supported. These data corroborate the results of an earlier investigation and add validity to the postulation of a generalized behavioral tendency to be impulsive (or reflective) in problem situations where the child should consider the validity of his answers. Impulsive children do not pause to consider the probable accuracy of their cognitive products, whether the situation involves visual matching of picture or reporting words in a recall situation.

The hypothesis that reflective *S*s are anxious over possible failure received only minimal support, and primarily among boys. The largest increase in intrusion errors and largest loss in recall score occurred among the reflective boys after they were told the lists they were about to hear were difficult. This inference requires the double assumption that the "threat" instruction is likely to generate worry over quality of performance for most of the children and that the sequellae of the anxiety are poor recall and more intrusions. Since there is no independent test of intensity of anxiety this nest of assumptions may be gratuitous. This conceptual problem is pervasive in psychology and indirect inference must often be used.

This interpretation is consistent with the data and current theory but is by no means the only way to interpret the findings.

As noted earlier, anxiety over the possibility of making a mistake is to be distinguished conceptually from expectancy of failure. A child may be so accustomed to failure that he has stopped protecting himself, and as a result, is not highly anxious over possible failure. It is likely that there is a curvilinear relationship between these two variables, with maximum values for the anxiety variable associated with moderate values of expectancy of failure. Children who are maximally or minimally confident of success are apt to be minimally anxious over possible failure.

The major implication of this work is to emphasize the significance of a conceptual tempo variable for cognitive products. Investigators working with "culturally deprived" children believe that one reason for their poor intellectual performance is their impulsive orientation.

The brain-damaged child, as well as the reading-retarded child, is more prone to be impulsive than reflective and his inferior intellectual performances are more often the result of impulsivity than inadequate verbal or knowledge resources. Therapeutic regimes for these children should consider the potential value of training reflection as a specific conceptual habit, independent of the specific substantive content of the material to be mastered.

REFERENCES

KAGAN, J. Impulsive and reflective children. In J. Krumboltz (Ed.), *Learning and the educational process*. Chicago: Rand, McNally, 1965. Pp. 133–161.

KAGAN, J., ROSMAN, B. C., DAY, D., ALBERT, J., & PHILLIPS, W. Information processing in the child. *Psychological Monographs*, 1964, **78**(1, Whole No. 578).

(Received December 4, 1964)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 25–29

PSYCHOPATHY AND CHOICE OF IMMEDIATE VERSUS DELAYED PUNISHMENT[1]

ROBERT D. HARE

University of British Columbia

This study tested the hypothesis that psychopathic Ss would show less preference for immediate shock than would nonpsychopathic Ss. 12 psychopathic and 12 nonpsychopathic criminals and 19 noncriminals were presented with 6 trials in which they were required to choose between an immediate shock and 1 delayed 10 sec. The results confirmed the hypothesis ($p < .001$). The psychopaths chose immediate shock 55.5% of the time while the nonpsychopaths chose it 82.3% of the time. The preference for immediate shock increased over trials for the nonpsychopathic Ss but not for the psychopathic ones. The results were interpreted in terms of a conditioned fear hypothesis.

Recent evidence indicates that normal subjects (*S*s) tend to prefer an immediate punishment (electric shock) over one that is delayed (Cook & Barnes, 1964), or randomly delayed (D'Amato & Gumenik, 1960). While it is obvious that many interacting variables are involved in choice behavior of this sort, relatively few empirical data concerning their effects are available. Among the variables likely to be of importance are (*a*) the relative delay, intensity, and probability of punishment associated with each of the alternatives presented to the *S*; (*b*) the type of punishment involved and the characteristics of the situation in which it is administered; and (*c*) certain characteristics (e.g., personality and motivational) of the individual making the choice. The present study investigated what was considered to be a particularly relevant example of the latter, namely, psychopathy.[2]

It is well known clinically that the psychopath tends to avoid immediate discomfort and that he appears to be relatively unconcerned about the long term consequences of his behavior (Cleckley, 1959; Karpman, 1961). On the basis of these and other considerations, it was expected that when faced with a choice between immediate and delayed shock, psychopaths would show less preference for immediate shock than would nonpsychopathic *S*s.

METHOD

The *S*s were all male volunteers and consisted of 12 psychopathic (Group P) and 12 nonpsychopathic (Group NP) inmates of the British Columbia Penitentiary, and 19 students (Group C) obtained from the university and a local adult education center. Mean age was 26.3 ($SD = 3.9$) for Group P, 27.8 ($SD = 3.6$) for Group NP, and 23.8 ($SD = 2.9$) for Group C. Mean number of years of formal education was 10.4 ($SD = 1.6$) for Group P, 10.2 ($SD = 1.7$) for Group NP, and 11.5 ($SD = 1.1$) for Group C. Mean Revised Beta IQ was 108.1 ($SD = 9.2$) for Group P, 106.9 ($SD = 8.9$) for Group NP, and 110.8 ($SD = 8.0$) for Group C. None of these differences was significant. Selection of the psychopaths was made by the institutional psychologists on the basis of criteria outlined by Cleckley (1959).

Testing was carried out in a quiet room in the appropriate institution. The *S* was seated before a horizontal panel upon which two telegraph keys were situated 12 inches apart. A small green pilot light was mounted at about eye-level on a vertical panel placed immediately behind the keys. A Model 1A Psychological Instruments Stimulator was used to deliver a brief electric shock via finger electrodes of the zinc, zinc-sulphate type (Lykken, 1959). The level of shock used was individually determined by gradually increasing the intensity until the *S* indicated that he was not prepared to accept anything stronger. The mean intensities of shock arrived at by this procedure and subsequently used in the experiment were 3.9, 3.6, and 3.8 milliamperes for Groups P, NP, and C, respectively. These intensities did not differ significantly from one other and are approximately the same as those obtained from similar *S*s in another study (Hare, in press).

---

[1] This research was supported by the President's Research Fund, U.B.C., and by Research Grant APA-139 from the National Research Council of Canada. Appreciation is extended to John Moloney, Warden of the B.C. Penitentiary and to S. A. Miller, Director of Research and Special Services for the Vancouver School Board, for making subjects and facilities available.

[2] The term refers to the entity described by Cleckley (1959) and to what Karpman (1961) terms the primary psychopath.

TABLE 1

TOTAL NUMBER AND PERCENTAGE OF CHOICES OF IMMEDIATE (0 SECS.) AND DELAYED (10 SECS.) SHOCK MADE BY EACH GROUP

Choice	Group					
	P		NP		C	
	0 secs.	10 secs.	0 secs.	10 secs.	0 secs.	10 secs.
Total	40	32	63	9	90	24
Percentage	55.5	44.5	87.5	12.5	78.9	21.1

TABLE 2

INDIVIDUAL CHOICES OF IMMEDIATE (1) AND DELAYED (D) SHOCK ON EACH TRIAL

	Trial						
Group P Subject	1	2	3	4	5	6	Σ1
1	D	1	D	1	D	1	3
2	1	1	1	1	1	1	6
3	1	1	1	D	1	D	4
4	D	D	D	D	D	D	0
5	D	1	1	D	1	D	3
6	1	1	D	1	1	1	5
7	1	D	1	1	D	1	4
8	D	D	1	D	1	1	3
9	1	D	1	D	1	D	3
10	1	1	1	1	1	1	6
11	D	1	D	1	D	1	3
12	D	D	D	D	D	D	0
Σ1	6	7	7	6	7	7	40
%1	50	58	58	50	58	58	55.5
Group NP Subject	1	2	3	4	5	6	Σ1
1	1	1	1	1	1	1	6
2	1	1	1	1	1	1	6
3	D	1	1	1	1	1	5
4	1	1	1	1	1	1	6
5	D	1	D	1	1	1	4
6	1	D	1	1	1	1	5
7	1	1	1	1	1	1	6
8	1	1	1	1	1	1	6
9	1	1	1	D	1	1	5
10	D	D	1	1	1	1	4
11	1	1	1	1	1	1	6
12	D	1	1	D	1	1	4
Σ1	8	10	11	10	12	12	63
%1	67	83	92	83	100	100	87.5
Group C Subject	1	2	3	4	5	6	Σ1
1	D	1	1	1	1	1	5
2	1	1	D	1	D	1	4
3	1	D	1	D	1	1	4
4	D	1	1	1	1	1	5
5	1	1	1	1	1	1	5
6	D	1	1	1	1	D	4
7	1	D	1	1	1	1	5
8	1	1	D	D	1	1	4
9	1	D	D	1	1	1	4
10	1	1	1	1	D	D	4
11	D	1	1	1	1	1	5
12	D	1	D	1	1	1	4
13	1	D	1	1	1	1	5
14	1	D	1	D	1	1	4
15	1	1	1	1	1	1	6
16	1	1	D	1	1	1	5
17	D	1	1	1	D	1	4
18	1	1	1	1	1	1	6
19	1	1	1	1	1	1	6
Σ1	13	14	14	16	16	17	90
%1	68	74	74	84	84	89	78.9

After the intensity of shock to be used had been determined, *S* was asked to place his dominant hand in front of the keys and midway between them. He was told that his task was simply to press one of the keys whenever the light came on, and that as a result he would receive a shock either immediately or after a 10-second delay, depending upon which key had been pressed. A card placed above each key informed *S* of the delay associated with the key. In order to familiarize him with the actual shocks delivered by the keys, he was required to press each key once. Six free-choice trials were then given, with about 30 seconds between each trial. The effect of possible position preferences was controlled by having the immediate shock associated with the left key for one-half of the *S*s and with the right key for the other half. After the experiment *S*s were briefly questioned about their experiences.

RESULTS AND DISCUSSION

The total number and percentage of choices of immediate and delayed shock made by each group are shown in Table 1. The mean number of trials (out of six) on which immediate shock was chosen was 3.33 for Group P, 5.25 for Group NP, and 4.74 for Group C. An analysis of variance indicated that these means differed significantly from one another, $F\ (2,40) = 7.93$, $p < .001$. Individual comparisons among these means (Winer, 1962, p. 100) revealed that Groups NP and C were not significantly different, $F\ (1,40) = 1.26$, $p > .10$, but that these two groups combined differed significantly from Group P, $F\ (1,40) = 15.86$, $p < .001$.

A more detailed impression of the choices made by each group can be gained from inspection of Table 2, which contains the choice of shock delay made by each *S* on each of the six trials. On the first trial, 67% of the nonpsychopathic *S*s chose immediate shock while only 50% of the psychopaths did so. Although in the expected direction, this difference was not significant. During the next five trials however, the number of nonpsychopathic *S*s who chose immediate shock increased significantly, Cochran's $Q = 9.7$, $df = 5$, $p < .05$ (Siegal, 1956), while the number of psychopaths doing so showed virtually no change.

The relatively strong overall preference for immediate shock shown by the nonpsychopathic *S*s (82.3% for Groups NP and C combined) is consistent with previous research involving both human (Cook & Barnes, 1964; D'Amato & Gumenik, 1960) and animal *S*s (Knapp, Kause & Perkins, 1959; Sidman & Boren, 1957), and also with the common experience that waiting for an unpleasant event to occur can be extremely distressing. The postexperimental comments made by these *S*s are pertinent here. Most stated that waiting for delayed shock produced a considerable amount of apprehension and that they wanted "to get it over with as soon as possible." The psychopaths, on the other hand, showed only a slight overall (55.5%) preference for immediate shock and reported that waiting for the occurrence of delayed shock bothered them very little.

A basis for interpreting the differential effects of delayed shock upon psychopathic and nonpsychopathic *S*s is provided by the theory that cues associated with painful stimulation acquire, through classical conditioning, the capacity to elicit fear responses (Miller, 1951). In normal persons, these cues would serve to elicit fear responses in the interval prior to anticipated pain or punishment. Presumably the aversive properties of this aroused fear (apprehension?) would summate with those of delayed shock and would be greater than the aversive properties of immediate shock alone. Since the aroused fear is a conditioned response, its magnitude and aversiveness would be expected to increase with repeated delayed shock experiences, and this would be reflected in an increased number of choices of immediate shock over trials. As Table 2 indicates, this is what happened in the case of the nonpsychopathic *S*s (Groups NP and C). The choice behavior of the psychopaths can be similarly interpreted. Thus, the finding that psychopaths may acquire conditioned fear responses slowly (Hare, 1965a; Lykken, 1957) suggests that, even after a number of delayed shock experiences, relatively little fear would be generated by cues preceding the shock. This would account for the failure of the psychopaths (Group P) to show an increased number of immediate shock choices over trials (see Table 2).

The results of this study, as well as the interpretation offered, are in accord with the recent finding that psychopathic *S*s showed less increase in palmar skin conductance in the interval prior to anticipated shock than did nonpsychopathic *S*s (Hare, 1965b; in press). This was taken to indicate that the emotional effects and the aversive properties of future pain or punishment are relatively small for the psychopath. Whereas the normal person finds it distressing to wait for some unpleasant event, the psychopath apparently does not.

A limitation of the present study is that it is not known to what extent nonsensory variables (Swets, 1961) affected the level of shock which *S*s reported to be their maximum. It is possible, for example, that some *S*s, in an attempt to avoid strong shock, reported that their maximum had been reached when in fact it had not. This possibility would particularly apply to the psychopaths, whose tendency to avoid discomfort is well known (Cleckley, 1959). Unfortunately, the method employed in this study did not allow the effects of such nonsensory variables to be evaluated directly. It will be recalled, however, that all three groups received approximately the same intensity of shock. This suggests that unless psychopaths have a pain threshold for shock that differs considerably from that of normal persons, the subjective intensity of shock was probably within the same range for all three groups. In this regard, Schachter and Latané (1964) found that psychopathic and nonpsychopathic *S*s did not differ significantly in their ratings of the degree of pain experienced when they were all administered the same level of shock.

While the present findings are consistent with the hypotheses that psychopaths lack the normal capacity to acquire conditioned fear responses and that the emotional significance of future punishment is of relatively little immediate concern to them, they may also be interpreted in another way. It is possible that the choice behavior of the psychopaths was determined not by the emotional consequences of delayed shock or by an inability to acquire fear responses, but by a deliberate attempt on their part to use the experimental situation to their own advantage. In the absence of any specific information about the purpose of the experiment, and knowing that it was being conducted by a psychologist who, among other things, might communicate the results to prison authorities, the psychopaths may have reasoned that any extreme or unusual pattern of responding would appear abnormal or revealing of themselves. As a result they may have intentionally employed a random or systematic distribution of choices. Thus, the advantage or payoff expected by the psychopaths for responding in a particular way (including their comments that waiting for delayed shock to occur did not bother them) may have outweighed whatever emotional effects delayed shock might have had. Inspection of Table 2 indicates that the pattern of choices of five *S*s (1, 5, 8, 9, 11) in Group P conforms to this alternative explanation of the results.

This interpretation raises the additional possibility that the results of other experimental studies of psychopathy, including those involving fear (GSR) conditioning (Hare, 1965a; Lykken, 1957), may have been similarly influenced by an attempt on the part of the *S*s to control the situation and to manipulate or "con" the experimenters. In other words, the psychopath may be able to control the amount of overt and autonomic emotionality displayed in a variety of social situations, perhaps because of a past history of reinforcement for such control and lack of display under similar circumstances. While this hypothesis is plausible and certainly coincides with the psychopath's well known penchant for manipulation, there is no direct experimental evidence to support it. On the other hand, the results of a number of recent studies involving avoidance learning (Lykken, 1957; Schachter & Latané, 1964), serial learning under conditions of punishment (Hetherington & Klinger, 1964), and the emotional effects of anticipated punishment (Hare, in press), are consistent with the hypothesis that psychopaths do not acquire fear responses as readily as do normal persons. As noted above, the present results are also consistent with this conditioned fear hypothesis. It is clear, however, that further research is necessary to determine the degree to which experimental findings, such as those obtained in the present study, are influenced by the psychopath's characteristic tendency to manipulate people and situations to meet his own needs. Research along these lines is currently being planned.

REFERENCES

CLECKLEY, H. M. Psychopathic states. In S. Arieti (Ed.), *American handbook of psychiatry.* Vol. 1. New York: Basic Books, 1959. Pp. 567–588.

COOK, J. O., & BARNES, L. W. Choice of delay of inevitable shock. *Journal of Abnormal and Social Psychology,* 1964, **68**, 669–672.

D'AMATO, M. R., & GUMENIK, W. E. Some effects of immediate versus randomly delayed shock on an instrumental response and cognitive processes. *Journal of Abnormal and Social Psychology,* 1960, **60**, 64–67.

HARE, R. D. Acquisition and generalization of a conditioned fear response in psychopathic and nonpsychopathic criminals. *Journal of Psychology,* 1965, **59**, 367–370. (a)

HARE, R. D. Psychopathy, fear arousal, and anticipated pain. *Psychological Reports,* 1965, **16**, 499–502. (b)

HARE, R. D. Temporal gradient of fear arousal in psychopaths. *Journal of Abnormal Psychology,* in press.

HETHERINGTON, E. M., & KLINGER, E. Psychopathy and punishment. *Journal of Abnormal and Social Psychology,* 1964, **69**, 113–115.

KARPMAN, B. The structure of neurosis: With special differentials between neurosis, psychosis, homosexuality, alcoholism, psychopathy, and criminality. *Archives of Criminal Psychodynamics,* 1961, **4**, 599–646.

KNAPP, R. K., KAUSE, R. H., & PERKINS, C. C. Immediate versus delayed shock in T-maze performance. *Journal of Experimental Psychology,* 1959, **58**, 357–362.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology,* 1957, **55**, 6–10.

LYKKEN, D. T. Properties of electrodes used in electrodermal measurement. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 629–634.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–472.

SCHACHTER, S., & LATANÉ, B. Crime, cognition and the autonomic nervous system. In D. Levine (Ed.), *Nebraska symposium on motivation: 1964*. Lincoln: Univer. Nebraska Press, 1964. Pp. 221–273.

SIDMAN, M., & BOREN, J. J. The relative aversiveness of warning signal and shock in an avoidance situation. *Journal of Abnormal and Social Psychology*, 1957, **55**, 339–344.

SIEGAL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SWETS, J. A. Is there a sensory threshold? *Science*, 1961, **134**, 168–177.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received December 31, 1964)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 30–41

EMPIRICAL CONTRIBUTIONS TO A THEORY OF AMBIVALENCE[1]

ABRAM MINKOWICH,[2] LINDA L. WEINGARTEN,[3] AND GERALD S. BLUM

University of Michigan

Ambivalence is defined as the coexistence of opposed emotional attitudes toward a significant figure in the social environment. A new technique for the assessment of ambivalence is described; its application in a series of studies on college students, hospitalized schizophrenics, and school children reviewed. The obtained relationships of ambivalence to sociocultural characteristics, perceived family relationships, psychosexual conflicts, defenses, and superego variables are presented. Data on the generality of ambivalent feelings toward parents, other authority figures, God, and siblings are also given. Finally, the empirical findings are pulled together to evolve a theoretical account of the psychodynamics of ambivalence, covering both etiological aspects and manifestations in the adult personality.

> "Even the normal individual feels, as it were, two souls in his breast; he fears an event and wishes it to come, as in the case of an operation or the acceptance of a new position. Such a double feeling tone exists most frequently and is particularly drastic when it concerns persons whom one hates or fears and at the same time loves. . . .
>
> But such ambivalent feeling tones are the exception with the normal person. On the whole he makes a decision from the contradictory values; he loves less because of accompanying bad qualities and hates less because of accompanying good qualities. But the abnormal person often cannot bring together these two tendencies; the hate and the love manifest themselves side by side without the two affects weakening or even influencing each other in any way . . . [Bleuler, 1951, p. 125]."

Ambivalence has been a popular concept since Bleuler introduced it 50 years ago. For the psychologist, however, its significance has been jeopardized by the vague, idiosyncratic, or inappropriate applications associated with its wide acceptance. Bleuler's use of the term was primarily descriptive, an abstraction from his observations of schizophrenic behavior. "Ambivalent complexes" were said to influence many forms of pathology and certain aspects of normal experience, such as dreams and creative acts. However, it was Freud and his followers who first ascribed to the construct widespread dynamic significance as a force in personality organization. Sources of ambivalent feelings, stages of their development, and manifestations at the individual, social, and cultural levels have all been treated at length.

Origins are traced to (*a*) the vagaries of instinctual expression, such as the polarities of love and hate, activity and passivity, introjection and projection; (*b*) problems common to the socialization of children everywhere, for example, contradictory needs for dependence and autonomy, or the dual role of socializing agents as protectors and disciplinarians; and (*c*) those specific environmental factors—particular family situations, child-rearing practices, and the like—which modify or enhance existing predispositions.

Stages in the development of ambivalence have been delineated from birth onward. Psychosexual conflicts during the early and late oral, anal, and phallic phases all are presumed to promote mixed feelings toward important

---

[1] Financial support for the studies reported in this paper was provided partly by a Rackham Fellowship to the first author; a United States Public Health Service Fellowship to the second; and National Institute of Mental Health Grant 1286 to the third.

[2] Now at Hebrew University, Jerusalem, Israel.

[3] Now at Mt. Sinai Hospital, New York City, New York.

figures in the child's environment. As in the case of other conflicts, defense mechanisms are brought into play, and vestiges remain in varying degree to mold subsequent personality formation during latency and adolescence. Intensification of ambivalence can contribute to psychopathology.

Manifestations of ambivalence at the intrapsychic level are exemplified by vacillations in self-esteem and inconsistencies in moral behavior. Socially, the extension of opposed attitudes toward the original objects, mother and father, to other figures outside the family can be observed. Finally, even at the cultural level, expressions of ambivalence are noted in religious rituals and taboos of many kinds.

The variety and complexity of referents mentioned above, compounded by an unfortunate tendency to equate ambivalence with conflict in general, weakens the clarity of the term in its current usage. We propose instead to limit the definition of ambivalence to the *co-existence of opposed emotional attitudes toward a significant figure in the social environment.*

PREVIOUS RESEARCH BEARING ON AMBIVALENCE

Since ambivalence has not been a clearly defined concept, a review of relevant research includes material from various areas—experimental, clinical, physiological, and social. The range of topics covers competing affects, response tendencies, social attitudes, and even neural interactions.

Investigation of competing affects has been pursued both by introspectionist and experimental approaches. One illustration is the work of Beebe-Center (1932), who used trained observers to report reactions to presentations of pleasant and unpleasant stimuli in simultaneous or successive pairings. He interpreted his findings as evidence for possible simultaneity at the conceptual but not at the affective, sensory level. Mendelsohn (1960) carried out an experiment in which hypnotically induced affects were pitted against each other simultaneously, and subjects' (*Ss*') reactions were recorded both in terms of verbal report and galvanic skin response (GSR). Anxiety decreased in the presence of contentment but not when accompanied by sadness or delight. Lanier (1941a, 1941b) studied the influence of affective conflict on memory, noting that words rated as "mixed" in emotional tone were recalled better, entailed longer judgment time, and showed heightened GSR.

Behavioral performance in the Lewinian and Hullian context of approach-avoidance bears some indirect relationship to ambivalence (e.g., Barker, 1942, 1946; Berlyne, 1957; Cartwright, 1941; Hovland & Sears, 1938). However, response tendencies rather than affects are of primary concern, and the conflict typically involves more than one object.

In the clinical-experimental tradition, Malmo and others (Malmo, Smith, & Kohlmeyer, 1956; Shagass & Malmo, 1954) found significant increases in muscular tension accompanying verbalization of material assessed by psychiatric methods as laden with ambivalent feelings. Osgood and Walker (1959) compared the language of suicide notes with ordinary letters, noting in the former more ambivalent "constructions" and "evaluative assertions." A series of studies with psychiatric patients has been carried out by Tarachow and others (Tarachow & Fink, 1953; Tarachow, Friedman & Korin, 1957; Tarachow, Friedman, & Korin, 1958), including interview and TAT comparisons between groups of neurotics; social history data; and perception and imagery tasks. Izard (1959) attempted to operationalize ambivalence as inconsistent behavior exhibited by paranoid schizophrenics (in contrast to normals) in the perception of photographs of human faces.

At the neurophysiological level, Roberts (1958a, 1958b) has demonstrated that both reward and punishment effects can be produced in the rat brain by stimulating an area with one electrode at a constant intensity level. Olds (1960) has obtained similar data suggesting so-called "ambivalence centers."

Studies in the social psychological realm have also dealt with a variety of topics. Seeman (1953) defined ambivalence as the subjective experience of role conflict in an investigation of attitudes of school superintendents and teachers toward the leadership role. Adorno, Frenkel-Brunswik, Levinson, and

Sanford (1950) observed that *S*s high in authoritarianism and ethnocentrism tended to handle ambivalent feelings by dichotomies and displacement, whereas similar feelings in other *S*s were "ego accepted" and more openly expressed toward original objects and "reality representatives." The Semantic Differential scale has been employed by Osgood (Osgood & Walker, 1959) to study attitudes toward political figures and issues in the 1952 presidential election. Findings were indicative of ambivalent perceptions of figures simultaneously admired and feared. Hofstätter (1956) used the same technique to investigate attitudes toward verbal concepts and colors. The opposites of "love" and "hate" turned out to be "loneliness" and "mildness," respectively. The color "red" loaded on two orthogonal factors and was interpreted as representing a true ambivalence relationship.

The preceding survey of empirical investigations indicates the heterogeneity of approaches to the measurement of ambivalence. The many divergent definitions and techniques have, for the most part, failed to clarify or systematically evaluate the concept. Most of the measures have suffered from an inability to differentiate ambivalence from general conflict, inconsistency, or behavioral response sets. Underlying affective predispositions toward psychologically meaningful objects have been tapped less frequently than competing responses in tasks designed to induce conflict. Clearly there is a need for methodological advance to encourage fruitful exploration of a topic so central to theories of personality formation. The development of a simple, easily administered technique for the assessment of individual differences in ambivalent feelings, generally applicable in studies of interpersonal relationships, is a worthwhile objective in itself.

DEVELOPMENT OF THE AMBIVALENCE ASSESSMENT TECHNIQUE (AAT)

Several considerations guided the initial formulation of a new and hopefully more relevant methodology to assess the coexistence of opposed emotional attitudes. On one hand, the desire to avoid pitfalls of clinical inference often inherent in projective and interview techniques dictated an objectively scorable approach. On the other, recognition of the vital part played by dynamic unconscious and defensive forces pointed to the necessity for subtle and indirect, rather than superficial modes of measurement. In addition, it was felt that a truly flexible and sensitive instrument should permit the evaluation of ambivalence toward a variety of figures in the social environment along a scale of affective intensity.

To accomplish these aims, an attempt was made to devise tasks varying in degree of structure and type of response. The facilitation of spontaneity by putting *S*s under time pressure; disguised presentation of stimuli and introduction of delays over time to allow expressions of inconsistency; and the utilization of clues to defensive operations, like response latency, all seemed promising attacks on the problem.

Minkowich (1960), as a part of an investigation of superego functions, devised the following procedures for tapping ambivalent feelings. Two paper-and-pencil tests were administered to 29 male and 37 female paid college students individually in two sessions, a week apart. In the first, *S* was presented with a booklet of 10 pages each headed by one of these figure names: Parent, Mother, Father, Brother, Sister, God, Judge, Male Teacher, Female Teacher, and Matronly Woman. He was instructed to write as quickly and frankly as possible the first 10 descriptive adjectives which he associated with each figure. The number and duration of response latencies was recorded by the examiner (*E*). After completing his associations to all 10 figures, *S* was asked to mark each of his responses in one of three ways: to assign a "+" if the adjective connoted a positive or good quality in the figure rated; a "−" if it seemed a negative or bad characteristic; or a "0" if neutral in connotation.

In the second session 1 week later, *S* evaluated the same 10 figures on a series of antonym pairs presented in the form of 7-point scales, as in Osgood's (Osgood, Suci, & Tannenbaum, 1957) Semantic Differential. Some equivalent pairs of antonyms were included, the criterion for equivalence being at least 85% agreement among a prior group of judges.[4] After this task, *S* reclassified each of his associations on the Adjective Booklet with his previous ratings of "+," "−," or "0" concealed.

From these procedures Minkowich (1960) derived 15 measures potentially relevant to ambivalence. These were subsequently intercorrelated and factor analyzed separately in four sets of data: male and female *S*s, each subdivided for ambivalence toward Mother and Father. All factor structures proved to be similar, including one replicated factor containing significant loadings of 10 of the measures. Ambivalence scores were then computed by summing *S*'s performance on the 10 variables.

Next, the technique was used by Weingarten (1962) with 28 hospitalized white male veterans diagnosed as chronic schizophrenics; and later with samples of 28 volunteer and 34 paid male undergraduates tested in group administrations. The schizophrenics' data again replicated the major ambivalence factor and yielded the additional information that composite scores based on a subset of 6 of the variables correlated very highly with scores from all 10. As a consequence, the following 6 variables were finally settled upon to score ambivalence:

1. Assignment of *both* positive and negative ratings (+/− ratio) to a figure on the first administration of the Adjective Booklet.
2. Rating *changes* of the same adjective between first and second administrations (+ one time, 0 or − the other, etc.).
3. Response latencies (blocking) in the writing of adjectives for a figure.
4. Inappropriate ratings of adjectives, for example, obviously negative attributes evaluated as "+."
5. Discrepancies in rating a figure on scales perceived to be equivalent in the Semantic Differential.
6. Restriction of Semantic Differential ratings to neutral midpoints of the scales.

EMPIRICAL FINDINGS

In the course of the series of studies mentioned in the preceding section many facets of ambivalence have been explored. Correlates of ambivalence toward parents were sought in sociocultural data, perceived family relationships, assessed psychosexual conflicts and defenses, and superego variables. Information on these topics was collected from (*a*) a biographical inventory dealing with perceived parental discipline, family background and composition, and religious practices; (*b*) a family attitudes questionnaire designed to elicit evaluations of parental behavior; and (*c*) administration of the Blacky Pictures, including modifications specifically tapping defense preferences and transgression fantasy, in addition to the standard assessment of psychosexual disturbance. The generality of ambivalent feelings across social figures, and control comparisons of ambivalence correlates with those of a hostility measure have also come under scrutiny.

Next are presented in turn the salient findings for the several college samples, the hospitalized schizophrenic group, and a sample of schoolboys. In the college groups the criterion of replication is employed for selecting results for presentation, that is, a finding is reported only if it occurred at a statistically significant level in two or more samples of *S*s (male or female) or in separate analyses of ambivalence toward both mother and father.

College Sample

Sociocultural correlates. Among the sociocultural variables in the male samples related to higher ambivalence scores toward parental figures, several concerned family religious affiliation and practices. Catholic *S*s were found to have more ambivalence toward Mother (AMB$_{MO}$). Male *S*s high in ambivalence toward both parents rated their religious sentiment as different in intensity from that of their fathers when asked to judge the strength of religious feeling of father, mother, and self ("strongly religious," "moderately religious," "hardly religious at all"). Also, *S*s obtaining high ambivalence scores toward father (AMB$_{FA}$) reported their own religious attendance as discrepant from that of their fathers. On this item, *S*s were required to rate church attendance on a 5-point scale ("regularly," "frequently," "sometimes," "rarely," "never"), and their judgments indicated dissimilarity between father and son in the frequency of religious observance.

With respect to parental occupation, the status level of working mothers was negatively related to AMB$_{MO}$ in males. The occupational status of each *S*'s mother was assigned a score in the following manner: no job outside home (0), manual labor (1), typist-clerk (2), secretary (3), nurse (4), teacher (5). Although none of the positions reported by *S*s for working mothers was in the executive or professional category, ambivalence scores toward the maternal figure were higher for *S*s reporting their mothers to have no outside job or low occupational status. AMB$_{MO}$ was higher for male *S*s who described their parents as having attributes which placed them in a "mass" rather than a "bureaucratic" setting. Persons were judged to belong to a "mass" socioeconomic level if any of the following conditions was applicable: (*a*) father employed as independent businessman, (*b*) father employed in a position involving no more than two higher levels of supervision, (*c*) either parent born in a foreign country, and (*d*) either parent born on a farm in the United States (Miller & Swanson, 1960).

---

[4] The following were judged to be equivalent scales: consistent-inconsistent and steady-vacillating, admired-despised and revered-disdained, sensible-foolish and wise-unwise, tolerant-intolerant and open minded-narrow minded, kind-unkind and considerate-inconsiderate.

Higher AMB$_{\text{FA}}$ scores were obtained by *S*s whose fathers had jobs which primarily concerned dealing in "things" rather than with "people" or "ideas." Occupations dealing in "things" included many technical fields, namely, dentistry, accounting, farming, merchandising, manufacturing, and brokerage. Jobs involving work with "people" were those of doctors, teachers, salesmen, personnel managers, and supervisors; while occupations categorized as dealing primarily with "ideas" included law, administration, research, art, architecture, and finance.

For female *S*s, ambivalence scores were related to two sociocultural variables: religious affiliation and socioeconomic setting. The *S*s whose parents were members of a religious minority (Catholics, Fundamentalists, or Jews) or who were themselves members of such a group obtained higher ambivalence scores toward both parents than females whose family religious affiliation was of a liberal Protestant denomination. In addition, more ambivalence toward parents was demonstrated by female *S*s from families with a "mass" socioeconomic background than by those coming from a "bureaucratic" setting.

Ambivalence and perceived family relationships. To explore the nature of family patterns which might affect the development of ambivalence toward parental figures, a family attitudes questionnaire was devised. The *S*s' perceptions of the childrearing attitudes and practices of each parent were tapped by means of forced-choice items pertaining to 12 family patterns. Each pattern of parental behavior was investigated by having *S*s judge statements dealing with (*a*) perceived underlying parental attitudes toward childrearing, (*b*) the child's behavior most likely to upset or irritate each of his parents, and (*c*) subjective criticisms of parental behavior. Two constellations from this questionnaire were related in two separate male samples to ambivalence toward both parents: paternal rejection and dominant wife. The *S*s ambivalent toward their parents endorsed items which described the father as too busy with other activities to be involved or interested in his children. They were also more likely than nonambivalent *S*s to describe the father as the less dominant parent who left the responsibilities of childrearing entirely to his wife.

Other data obtained from the biographical questionnaire showed a variety of relationships between ambivalent feelings toward parents and parental disciplinary practices. For instance, corporal punishment from the maternal figure was frequently reported by male *S*s high on AMB$_{\text{MO}}$. The *S*s with greater ambivalence toward the father also reported corporal punishment from the paternal figure more often than *S*s with lower AMB$_{\text{FA}}$ scores. In the female college sample the same result appeared, that is, reports of parental corporal punishment from *S*s high on the ambivalence measure for both parents. Corporal punishment was scored when *S*s indicated spanking and slapping as a frequently employed form of punishment.

Rewards and punishments from each parent were rated by *S*s on separate 5-point scales of frequency ("never," "rarely," "sometimes," "often," and "very frequently"), and these ratings were also compared with each other to arrive at a ratio of rewards and punishments from both mother and father. Male *S*s obtaining higher AMB$_{\text{MO}}$ scores were found to report (*a*) less reward and (*b*) more punishment than reward from the maternal figure. The *S*s with more AMB$_{\text{FA}}$ reported less frequent reward from mother and father and described the paternal figure as giving lesser or equal amounts of reward in comparison to punishment.

Female *S*s with elevated AMB$_{\text{MO}}$ scores checked infrequent reward from the maternal figure whereas punishment was perceived to have been meted out with equal or greater frequency by the mothers. Similarly, high AMB$_{\text{FA}}$ covaried with subjective reports of punishment exceeding or equal to frequency of reward from the father.

Inconsistency in maternal disciplinary behavior was correlated with higher AMB$_{\text{MO}}$ in male *S*s. Inconsistency of discipline patterns was rated on a 4-point scale ranging from "always consistent" to "generally inconsistent." Maternal noninvolvement in discipline was also related positively to AMB$_{\text{MO}}$. Noninvolvement was determined by a composite score based on a variety of other questionnaire items such as infrequent reward, and taking the child's behavior for granted when punishment or reward might have been expected.

A tendency was observed for male *S*s, high in ambivalence toward the father, to remember less praise for proper behavior from the maternal figure than did *S*s with lower AMB$_{\text{FA}}$. Conversely, *S*s obtaining greater AMB$_{\text{MO}}$ scores more often described their fathers as failing to praise and encourage. The latter refers to an item from a 4-part question dealing with the type of reaction *S*s observed in their parents in response to their own good behavior. Other items included material reward, taking good behavior for granted, and minimization of the significance of a good deed. Male *S*s with high AMB$_{\text{MO}}$ and AMB$_{\text{FA}}$ reported a tendency on the part of the maternal figure to take good behavior for granted. Those female *S*s ambivalent toward both parents also indicated minimal rewards from mother and father in response to good behavior.

With respect to parental punitive practices, seven items were listed and *S*s were asked to rank order the three most likely forms of punishment. The *S*s ambivalent toward the father were more likely to indicate spanking and slapping as a characteristic paternal response and less likely to describe the father as reasoning or explaining the child's mistake. Similarly, high AMB$_{\text{MO}}$ scores covaried with male *S*s' reports of physical and verbal punishment (spanking and scolding) from the mother and were negatively related to maternal reactions of reasoning and explanation. Less reasoning on the part of Mother was observed in relation to high AMB$_{FA}$, too. The *Ss* who reported that the maternal figure teased, made fun of, or ridiculed a bad child also tended to show more ambivalence toward Mother.

A question pertaining to the childrearing objectives favored by *Ss*, proved to have several relationships to ambivalence scores. Male *Ss* scoring high in ambivalent conflicts toward either parent stressed the importance of allowing children freedom to decide things for themselves. Such *Ss* choose less frequently than low scoring *Ss* alternatives concerning the inculcation of moral values ("stress the importance of moral standards" and "bring the child up to be honest and reliable"). Other extremely frequent choices for most *Ss* were items concerned with love and guidance. AMB$_{MO}$ was negatively related to the choice of these items.

Conflicts, defenses, and ambivalence. Independent measures of psychosexual conflict were obtained from a group administration of the Blacky Pictures to the Weingarten samples. Following the revised scoring system for research use (Blum, 1962), factor scores were derived from *Ss*' responses to the stimuli. AMB$_{MO}$ covaried with *Oral Craving,* a factor describing Blacky's (B's) voracious craving for oral supplies and concern over potential deprivation. Elevated ambivalence scores toward both parents were manifested by male *Ss* high on the factor *Exploitation,* which deals primarily with B's selfish expression of aggressive impulses, especially anal exploitation of parental figures. In addition, *Undisguised Oedipal Involvement,* a factor saturated with various sexual references toward Mama (M), comments about B's jealousy toward Papa (P) and expected retaliation, was significantly related to heightened ambivalence toward both parents.

Characteristic defensive operations in dealing with psychosexual conflict were investigated by means of the Blacky Defense Preference Inquiry (Blum, 1956). The *Ss* showing high AMB$_{MO}$ were found to employ "expressive" defenses such as projection, regression, and intellectualization more frequently than *Ss* obtaining lower scores. The *Ss* with greater conflict toward both parents tended not to choose generally "avoidant" defenses such as avoidance and reaction formation, especially when dealing with stimuli involving oral aggression, anal sadism, oedipal intensity, and masturbation guilt.

Females demonstrated more specific relationships between ambivalence and defense choice. To the oral aggression card showing B chewing M's collar, ambivalent females chose the projection and intellectualization alternatives more frequently and the avoidance and reaction formation items less frequently than *Ss* obtaining lower AMB$_{MO}$ and AMB$_{FA}$ scores. The *Ss* high in ambivalent conflict toward the mother were found to select the regression alternative in response to the card depicting the oedipal situation. In relation to the Blacky sibling rivalry scene, *Ss* with higher AMB$_{MO}$ chose avoidance and reaction formation less frequently. The *Ss* with higher AMB$_{FA}$ scores tended to prefer avoidance as a defense in response to the anal sadism card.

Superego variables and ambivalence. In psychoanalytic theory ambivalence is closely related to the formation and functioning of the superego. The internalization of moral standards by identification with parents is assumed to contribute to the solution of the child's ambivalence toward them. According to these notions, individuals with well-internalized moral standards should be expected to demonstrate less ambivalence toward their parents than those with superego defects.

This hypothesis was tested in the Minkowich male and female samples. Superego functioning was assessed from *Ss*' stories on six Blacky pictures (II, III, IV, V, VI, and VIII) and their responses to a specially modified inquiry for these pictures. Stories and inquiry responses were scored for three categories of fantasy about behavior: compliance, transgression, and noncommittal content. Other items in the specially devised inquiry pertained to four categories of affect: (*a*) an *internal orientation* consisting of satisfaction over self-control and guilt over transgression, (*b*) an *external orientation* including anticipation of rewards for compliance and fear of punishment for transgression, (*c*) an *impulsive orientation* consisting of frustration over impulse inhibition and also satisfaction from impulse gratification, and (*d*) an *avoidant orientation* derived from noncommittal responses to questions about affect resulting from compliance or transgression. An additional affect category was based on the inquiry to Picture IX (Guilt Feelings) which contained items of an unstructured nature, thus providing a score for diffuse or free-floating guilt.

In conformity with the underlying hypothesis, both males and females demonstrated positive relationships between ambivalence toward parents and composite scores of transgression fantasy based on all six Blacky pictures. Meaningful relationships were also found between ambivalence and affect scores. *Internal orientation* correlated negatively with ambivalence toward both parents in the case of females and with AMB$_{MO}$ for males. The male data revealed a significantly positive relationship between ambivalence toward both parents and free-floating unstructured guilt. For the *external orientation* the direction of relationships was positive in both samples but stronger for females. Males with greater AMB$_{MO}$ were high on both components of the *impulsive orientation*—frustration from impulse inhibition and satisfaction over impulse gratification. For females, ambivalence toward both parents correlated significantly with the first component.

Generality of ambivalence. The Minkowich data for male and female college samples were explored by the writers for evidence of the generality of ambivalence in attitudes toward parents, parental surrogates, and siblings. Basic to psychoanalytic theory is the assumption that attitudes toward parents shape attitudes toward other figures in the social environment. Generalization from childhood experiences is given special emphasis by Freud in his dis-

TABLE 1

INTERCORRELATIONS OF MALE AMBIVALENCE SCORES AMONG SEVEN FIGURES

Figure	1 Parent	2 Father	3 Mother	4 Brother	5 Sister	6 God	7 Judge
1		30	50**	02	06	04	13
2			25	06	−04	26	08
3				24	18	28	16
4					36*	39**	04
5						20	00
6							09
7							

Note.—Minkowich sample $N = 29$.
$^{*} p < .05$.
$^{**} p < .01$.

cussions of character structure and, in particular, ambivalence. All social attitudes are said to be inevitably invested with some degree of ambivalence as a consequence both of phylogenetic factors and of unavoidable contradictions in parent-child relationships. Even religious attitudes are ascribed to these molding forces:

> God is a father substitute . . . a reproduction of the father as seen and met with in childhood—as the individual sees him in his own childhood and as mankind saw him in prehistoric times in the father of the primal horde. . . . The attitudes contained in the original idea of the nature of God are but a reflection of the ambivalence governing the relations of an individual to his personal father [Freud, 1948, Vol. IV, p. 449].

To test this assumption of generality, ambivalence scores for the various stimulus figures were intercorrelated.

Table 1 presents the matrix for males of ambivalence scores among seven figures. Although all but two *r*'s are in the positive direction, the majority are low. Three reached or exceeded the 5% level of significance (one to be expected by chance). The corresponding matrix for females, given in Table 2, clearly contains more positive interrelationships. Twelve of the 21 are significant beyond the 5% level and 6 others approach significance.

Next, each matrix was factor analyzed by the varimax method with four orthogonal factors extracted for both sexes. Three of the male factors were highly similar to corresponding female ones. One pair is interpretable as general ambivalence toward parental figures, primarily Mother and, to a lesser extent, Father. It is interesting that Sister also loads on this factor in the female group. The second pair reflects sibling ambivalence, with Brother-Sister loadings somewhat higher for females. The third is identifiable as a Father-God ambivalence factor, augmented in females by Judge, thereby lending empirical support to Freud's comment quoted above.

Schizophrenic Sample

Ambivalence scores from the group of schizophrenic males, patients on the maximum privileges ward of a large community hospital who had been selected by ward physicians as cooperative and free of complicating organic illness, were related to responses to a personal history questionnaire. From the latter instrument, 46 variables were derived, dealing with demographic background, parental disciplinary practices, *Ss*' preferred childrearing objectives, and information about behavior patterns, and illnesses. Eight orthogonal factors emerged from a varimax rotation procedure, and on five the ambivalence measures showed appreciable loadings.

The five factors were interpreted in the following manner. Factor I, primarily a sociocultural dimen-

TABLE 2

INTERCORRELATIONS OF FEMALE AMBIVALENCE SCORES AMONG SEVEN FIGURES

Figure	1 Parent	2 Father	3 Mother	4 Brother	5 Sister	6 God	7 Judge
1		72**	63**	37*	27	34*	35*
2			57**	37*	05	41**	67**
3				30	35*	27	24
4					58**	26	09
5						28	09
6							33*
7							

Note.—Minkowich sample $N = 37$.
$^{*} p < .05$.
$^{**} p < .01$.

sion, links AMB$_{FA}$ and hostility toward Mother with a relatively poor social and economic position. Parents—often foreign born, members of minority religious faiths, and large city dwellers—are perceived as less successful in material accomplishments. The mother is described as the chief disciplinarian notwithstanding the father's reported benevolence. High-scoring *S*s on this factor are apt to reject emphasis on a child's achievement and instead prefer the option of allowing children a great deal of personal freedom.

In Factor II, a broken home or disturbed parental marriage, the presence of foster parents, and punitive treatment by the father or paternal substitute all go with AMB$_{FA}$. The father is a nonrewarding figure, often harsh in handling misbehavior, and the real mother may be dead. Rejecting an emphasis on inculcating morality, *S*s with this background also endorse a parent's granting children freedom to decide things for themselves.

The third factor describes a relationship between AMB$_{MO}$ and behaviors suggestive of maternal apathy or neglect. Childhood residence in an urban community and few or no siblings are additional background features associated with this constellation.

Factor IV, on which AMB$_{MO}$ also loads, deals with paternal disciplinary practices, especially father's reported failure to reward, praise, punish, or be involved in childrearing. The family, relatively prosperous in a rural community, is apt to include many children, whom the mother controls by corporal punishment. High scoring *S*s on this factor are likely to choose social achievement and omit morality as important values to instill in children.

Family religious practices are a major focus of Factor V. Less AMB$_{MO}$ is expressed by older respondents who report strict religious training and a family pattern of frequent church attendance. From larger families and unbroken homes, high-scoring *S*s are apt to perceive the mother as a rewarding parent. Obviously the factor may be described alternatively as a relationship between higher AMB$_{MO}$ scores and the absence of these background characteristics.

Schoolboy Sample

Gallatin (1962), employing an adaptation of the Ambivalence Assessment Technique for younger *S*s, administered a shortened and simplified version to 29 male sixth, seventh, and eighth graders in two sessions. On the first occasion *S*, seen individually, provided six associations to each of five figures (Mother, Father, Self, Sister, and Brother) as *E* recorded the response latencies. The *S* then rated each adjective with a "+," "−," or "0." During the second session approximately 3 weeks later, small groups of *S*s rerated their associations and completed for the same five figures revised Semantic Differential scales (reduced in number from 20 to 15 antonym pairs and simplified for the level of verbal comprehension). Ambivalence scores were computed for the two parental figures. These scores were compared with each other and with measures of perceived dissimilarity to parents, agreement with parents, home atmosphere, and self-ambivalence.

TABLE 3

GENERAL CORRELATES OF AMBIVALENCE TOWARD BOTH PARENTS (ADULT MALES)

Religious practice:	S discrepant from Fa in strength of religious feelings
Parental discipline:	Mo took good behavior for granted Mo did not reason with S or try to explain his mistakes
Family pattern:	Paternal rejection Mo dominant parent in childrearing
Ss' own childrearing objectives:	Allow child freedom to decide things himself, rather than inculcating morality
Blacky factors:	Anal exploitation Undisguised oedipal involvement
Defense preference:	Choice of expressive defenses, especially projection, rather than avoidant ones
Superego variables:	Transgression fantasy Free-floating guilt

While perceived dissimilarity, a measure of identification based on dissimilar descriptions of self and parental figures on the Semantic Differential, was positively related to ambivalence, the result was significant only with respect to the AMB$_{FA}$ scores. In answering multiple-choice questions about childrearing decisions, ambivalent *S*s did not show greater disagreement with their parents, who also filled out the forms, than did low scoring *S*s. There was a tendency for ambivalent *S*s to come from home atmospheres rated as "conflicted" on the basis of material from parent-teacher conferences. A strong relationship was also observed between ambivalence scores toward Self and toward Father.

Summary of Findings

Tables 3, 4, and 5 summarize the major results, already described, which bear on ambivalence toward parents in the several male samples and one female sample.

DISCUSSION

Methodological Issues

Since theory leans heavily on the method by which it is tested, we must answer several pertinent questions concerning the measurement of ambivalence before drawing theoretical inferences from the empirical data. The first can be stated simply: Is the ambivalence test measuring something reliably? Incon-

TABLE 4

SPECIFIC CORRELATES OF AMBIVALENCE TOWARD EACH PARENT (ADULT MALES)

	AMB$_{\mathrm{MO}}$	AMB$_{\mathrm{FA}}$
Demographic characteristics:	Mass setting	
Religion:	Catholic	
Occupation:	Mother's job of low social status	Father's job deals with things, not ideas or people
Discipline:	Corporal punishment (Mo) Punishment equal to or exceeding reward (Mo) Inconsistency in Mother's discipline Noninvolvement in discipline (Mo) Mother teased, made fun of, ridiculed child	Corporal punishment (Fa) Punishment equal to or exceeding reward (Fa)
Religious practices:		S discrepant from Fa in frequency of religious observance
Blacky factors:	Oral craving	
Superego variables:	Impulse orientation Deny guilt in specific contexts	

sistent behavior on the test conceivably could reflect random responding rather than meaningful vacillation. Also, response sets on the part of *S*s could be operating, such as bias in favor of choosing socially desirable alternatives.

A number of lines of evidence converge to rule out both these possibilities. Those ambivalence subscores which are not based upon inconsistent behavior (latency and scale restriction) correlate positively with the other four. Random responding is rendered still more unlikely by the consistency of relationships between test scores and the wide variety of dependent variables measured in diverse samples of *S*s. The factor of social desirability is negated by the differential results obtained with ambivalence and hostility scores [5] derived from the same raw data and

TABLE 5

GENERAL AND SPECIFIC CORRELATES OF AMBIVALENCE TOWARD PARENTS (ADULT FEMALES)

A. General correlates of ambivalence toward parents

Demographic characteristics:	Mass setting
Religion:	Minority religious affiliation
Discipline:	Corporal punishment by either parent Good behavior not rewarded by either parent
Defense preferences:	Projection or intellectualization, rather than avoidance or reaction formation chosen for II (Oral aggression)
Superego variables:	Transgression fantasy Fear of punishment Unhappy over impulse inhibition Deny guilt in specific contexts

B. Specific correlates of ambivalence toward each parent

	AMB$_{\mathrm{MO}}$	AMB$_{\mathrm{FA}}$
Discipline:	Reward less than punishment (Father)	Reward less than punishment (Father)
Defense preferences:	Regression on IV (Oedipal involvement) *Not* avoidance or reaction formation on VIII (Sibling rivalry)	*Not* avoidance on III (Anal aggression)

---

[5] In an attempt to clarify the obviously overlapping relationship between ambivalence and hostility (the latter constituting a major component of coexisting, opposed emotional attitudes represented in the former), an index of hostility was computed from the following responses in the Ambivalence Assess-

equally susceptible to response set. For example, AMB_{MO} in males correlates significantly with Blacky Factor IV-A (Undisguised Oedipal Involvement) whereas hostility does not. Similarly, transgression fantasy in two Blacky areas is related to ambivalence but not to hostility. Significant correlations of ambivalence scores with dependent variables where social desirability does not play a role—for example, DPI items constructed to have roughly equal popularity—also point away from a response-set interpretation. In general, the consistent, differentiated patterns of obtained results are not compatible with gross assumptions of casual or biased test-taking attitudes.

A second question: Is it reasonable to place so much weight upon *S*'s currently reported perceptions in the absence of direct observation of his actual socialization experiences? Essentially the techniques are tapping cumulated remnants of the life history. Here it can be argued that such perceptions may be the best indicators of ambivalence, since they portray *S*'s incorporated attitudes molded by personal experience. At all stages the perceived world is a more influential force in personality development than the objectively "real" one, and those vestiges which remain active in *S*'s current recall are more likely to reflect crucial subjective experiences from the past.

But willingness to concede the adequacy of measurement methods and even the validity of data derived from personal reports still leaves unanswered a further question. Can inferences concerning the childhood *etiology* of ambivalence be drawn legitimately from current correlational assessment of young adults? Although it is important to note the nature of the present concomitants of ambivalence in the structure of personality, intriguing aspects of the problem also have to do with its origins. The selection of materials for inclusion in the test battery reflects these interests. Factual information presumed to relate to the socialization process, such as religious affiliation and type of family setting, is included along with indirect, projective measurement of psychosexual conflicts, defenses, morality, etc., as well as *S*'s purported recollection of actual childhood events. This multifaceted approach enables etiological inferences to be drawn with a fair degree of assurance, probably somewhere between the limits of speculation based upon unsystematic clinical observation at one end and the "true" state of affairs at the other.

Notes on the Etiology of Ambivalence

Certain sociocultural and perceived family relationships recur consistently in the empirical findings. High ambivalence is associated with minority religious affiliation, mass setting, and additional attributes which suggest relatively lower socioeconomic status in parental background. More punitive or apathetic discipline, including corporal punishment, failure to reward good behavior, maternal noninvolvement, and inconsistency are also reported by ambivalent *S*s. It may be that such families are apt to use harsher modes of discipline as a consequence of their personal frustration in coping with difficult environmental circumstances. A contributing factor may be membership in a minority religious group, for studies (Miller & Swanson, 1958) have suggested that severe discipline sometimes stems specifically from certain religious convictions.

Another alternative is that parents with socially less desirable characteristics do not serve as satisfactory objects for identification initially and even later, thereby perpetuating the ambivalence conflict. Those family patterns in which the father is perceived as rejecting, or else dominated by his wife, can lead to identification difficulties for male *S*s. The schoolboy data, in supporting home conflict as a feature of family background in ambivalent male *S*s, provide additional support for this interpretation.

Personality Correlates of Ambivalence

Evidence of actual identification disturbances is suggested by the more ambivalent

---

ment Technique: negative or noncommittal ratings for the figures in the Adjective Booklet; and negative evaluations on the Semantic Differential scales. The median intercorrelation between ambivalence and hostility for each parent across the various samples is +.63. This finding suggests that, despite some built-in overlap, a large portion of the variance of the ambivalence scores cannot be accounted for in terms of hostility.

schoolboys' tendency to perceive themselves as different from their fathers on evaluative dimensions. Later in life these trends continue to appear as personality correlates, for ambivalent college students report patterns in church attendance and strength of religious feeling which are discrepant from those of their parents.

A relationship between ambivalence and moral behavior, suggested by psychoanalytic theory, is demonstrated in the present data. Ambivalent individuals display defects in the internalization of moral standards, in fantasied violations of sexual and aggressive norms, and in minimizing feelings of remorse over deviant behavior. Moral decisions and feelings appear to be based on external factors or impulsive tendencies rather than on introjected values. Accompanying this rebellious attitude, however, is an undifferentiated general guilt reaction. The rebellion is reflected further in ambivalent *S*s' emphasis on freedom as a childrearing objective rather than on love, guidance, or the inculcation of moral standards. Resentment of parents and revolt against their value systems are symptoms of underlying conflict.

The dynamic picture suggested by personality correlates of high ambivalence is one of unresolved oedipal conflict and exploitative impulses directed toward parental figures. Not only do ambivalent children fail to incorporate parental models, but early patterns of psychosexual conflict remain dominant aspects of their personality functioning. Oral conflict, particularly unmet dependency needs, is typical of *S*s whose opposed emotional attitudes center around the mother. Further support for analytic propositions is obtained from the generality study which demonstrates the subsequent extension or transfer of coexisting opposed feelings to persons outside the immediate family. Emerging factors seem to follow well-defined patterns centering around male authority figures, female authority figures, and peers. The use of particular defenses, especially projection, is also typical for ambivalent *S*s when dealing with psychosexual conflict, but further investigation seems necessary to establish the significance of ambivalence in this and other areas relating to psychopathology.

The above remarks concerning the etiology and correlates of ambivalence indicate support for some very generally formulated notions of psychoanalysis. But the real need is to evolve a detailed, well-articulated theoretical account of the origins and manifestations of this crucial personality characteristic. The data acquired in the series of studies reported here represent a good beginning in that direction. There are many leads to follow up and a promising technique, of demonstrated value, has been developed to pursue them.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.

Barker, R. G. An experimental study of the resolution of conflict by children. In Q. McNemar & Maud A. Merrill (Eds.), *Studies in personality.* New York: McGraw-Hill, 1942. Pp. 13–34.

Barker, R. G. An experimental study of the relationship between certainty of choice and the relative valence of the alternatives. *Journal of Personality,* 1946, **15**, 41–52.

Beebe-Center, J. G. *Pleasantness and unpleasantness.* New York: Van Nostrand, 1932.

Berlyne, D. E. Conflict and choice time. *British Journal of Psychology,* 1957, **48**, 106–118.

Bleuler, E. P. *Textbook of psychiatry.* New York: Dover, 1951.

Blum, G. S. Defense preferences in four countries. *Journal of Projective Techniques,* 1956, **20**, 33–41.

Blum, G. S. A guide for research use of the Blacky pictures. *Journal of Projective Techniques,* 1962, **26**, 3–29.

Cartwright, D. Relation of decision-time to the categories of response. *American Journal of Psychology,* 1941, **54**, 174–196.

Freud, S. *Collected papers of Sigmund Freud.* Vol. IV. London: Hogarth, 1948.

Gallatin, Judith. Some correlates of ambivalence toward parents. Unpublished undergraduate honors dissertation, University of Michigan, 1962.

Hofstätter, P. R. Farbsymobolik und ambivalenz (Color symbolism and ambivalence). *Psychologische Beiträge,* 1956, **2**, 526–539.

Hovland, C. I., & Sears, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *Journal of Experimental Psychology,* 1938, **23**, 477–493.

Izard, C. E. Paranoid schizophrenic and normal subjects' perceptions of photographs of human faces. *Journal of Consulting Psychology,* 1959, **23**, 119–124.

Lanier, L. H. An experimental study of "affective conflict." *Journal of Psychology,* 1941, **11**, 199–217. (a)

LANIER, L. H. Incidental memory for words of differing affective value. *Journal of Psychology*, 1941, **11**, 219–228. (b)

MALMO, R. B., SMITH, A. A., & KOHLMEYER, W. A. Motor manifestation of conflict in interview: A case study. *Journal of Abnormal and Social Psychology*, 1956, **52**, 268–271.

MENDELSOHN, G. A. Experiments on the psychoanalytic mechanism of isolation. Unpublished doctoral dissertation, University of Michigan, 1960.

MILLER, D. R., & SWANSON, G. E. *The changing American parent*. New York: Wiley, 1958.

MILLER, D. R., & SWANSON, G. E. *Inner conflict and defense*. New York: Holt-Dryden, 1960.

MINKOWICH, A. Correlates of superego functions. Unpublished doctoral dissertation, University of Michigan, 1960.

OLDS, J. Approach-avoidance dissociation in rat brain. *American Journal of Physiology*, 1960, **199**, 965–968.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

OSGOOD, C. E., & WALKER, EVELYN G. Motivation and language behavior: A content analysis of suicide notes. *Journal of Abnormal and Social Psychology*, 1959, **59**, 58–67.

ROBERTS, W. W. Both rewarding and punishing effects from stimulation of posterior hypothalamus with same electrode at same intensity. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 400–407. (a)

ROBERTS, W. W. Rapid escape learning without avoidance learning motivated by hypothalamic stimulation in cats. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 391–399. (b)

SEEMAN, M. Role conflict and ambivalence in leadership. *American Sociological Review*, 1953, **18**, 373–380.

SHAGASS, C., & MALMO, R. B. Psychodynamic themes and localized muscular tension during psychotherapy. *Psychosomatic Medicine*, 1954, **16**, 295–313.

TARACHOW, S., & FINK, M. Absence of a parent as a specific factor determining choice of neurosis. *Journal of Hillside Hospital*, 1953, **2**, 67–71.

TARACHOW, S., FRIEDMAN, S., & KORIN, H. Perceptual experiments in a study of ambivalence. *American Medical Association Archives of Neurology and Psychiatry*, 1957, **78**, 167–176.

TARACHOW, S., FRIEDMAN, S., & KORIN, H. Studies in ambivalence. *Journal of Hillside Hospital*, 1958, **7**, 67–97.

WEINGARTEN, L. L. Correlates of ambivalence toward parental figures. Unpublished doctoral dissertation, University of Michigan, 1962.

(Received February 15, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 42–51

HYPNOTIC BEHAVIOR:

THE DIFFERENTIATION OF TRANCE AND DEMAND CHARACTERISTIC VARIABLES[1]

KENNETH BOWERS[2]

University of Illinois

Hypnotic Ss received posthypnotic suggestions (a) to begin all sentences to the ostensible experimental (Taffel) task with "he" and "they," and (b) to be unaware of and amnesic for this fact. Waking simulator Ss received identical suggestions preceded by instructions to behave later as though they had been hypnotized when they received the suggestions. During a postexperimental inquiry with a different E, 8 of 14 hypnotic Ss were amnesic for their experimental behavior; none of the 13 simulating Ss were amnesic ($p < .0005$). All simulating Ss testified that their use of "he" and "they" was voluntary; the reverse was true for 12 of 14 hypnotic Ss ($p < .00005$). The results of this experiment demonstrate that hypnotic behavior is not wholly reducible to acting in accordance with demand characteristics, and suggest that there is a "state" of hypnosis within which suggestions have a peculiarly potent effect.

Most investigators interested in hypnosis believe that there is an hypnotic state which fundamentally differs from the waking state (Ås, 1963; Erickson, 1944, 1952; Marcuse, 1959; Orne, 1959; Shor, 1960, Weitzenhoffer, 1953, 1957, 1963; White, 1941). If there is a state of hypnosis, it would seem that there ought to be manifestations unique to it; yet nearly all of its purported behavioral indices have turned out to be spurious and/or evincible by procedures other than those traditionally termed hypnotic (Orne, 1959; Barber & Calverley, 1962, 1964a; Barber & Hahn, 1962; London & Fuhrer, 1961).

It is the contention of the present investigator that the search for behavior unique to the hypnotic state is understandable but somewhat misguided, and furthermore, that unequivocal demonstration of the existence of an hypnotic state is in no way contingent upon a demonstration that certain behaviors are unique to it. It may well be that almost any behavior, however bizarre, that is elicited from a somnambulist can be duplicated by a waking person providing only that (*a*) he is privy to information concerning how and in what manner he is supposed to behave, and (*b*) the stakes for a successful performance are high enough.

Nevertheless, it has been suggested that equivalent behavior obtained from simulator and hypnotic subjects is a sufficient basis for inferring an equivalence in their state, that this inferred state-equivalence casts doubt upon the reality of the hypnotic as distinct from the waking, motivated state, and that therefore "the concepts of 'trance' and 'hypnosis' may no longer be useful . . ." [Barber & Calverley, 1962, p. 388]. It is here claimed that any inference to equivalent states from equivalent behavior must ultimately devolve upon the *conditions under which similarities of behavior are obtained* (Weitzenhoffer, 1963). It is moreover claimed that the control conditions of many previous hypnotic experiments demonstrating no difference between hypnotic and simulating subjects disallow any such inference.

The effect of demand characteristics and expectancies (by whatever name) on hypnotic behavior seems to be widely acknowledged (Erickson, 1944; Orne, 1959; Sidis, 1906), but their effect on the behavior of simulating controls has not been equally appreciated. As a result, simulators are often differentially exposed to procedures which inform them of what constitutes "hypnotic" behavior and/or which communicates the investigator's expectancy that the simulator should behave in a prescribed, hypnoticlike fashion in order to make the experiment a success (e.g., Barber & Calverley, 1963, Experiment I; Orne, 1959). These kinds of instructions are generally termed motivating instructions—and no doubt they do enhance motivation; but they do so in part by altering the experimental demand characteristics and the subject's expectancies of how he is to behave. Most subjects in an experiment will ordinarily behave in a way commensurate with what they feel is expected of them (Orne, 1959, 1962); to explain motivated control subjects' behavior with reference to motivation only thus seems to be misleading and inaccurate. By contrast, hypnotic subjects are often told just to let happen whatever happens without attempting to help or hinder the operator. Similar behavior elicited from subjects in each of these two groups could quite obviously occur for two different reasons. The present investigator submits that selectively providing only the simulating subjects with expectancies eventuating in behavior comparable to that evinced by hypnotic subjects confounds the effects of trance variables on one hand, and demand characteristics and expectancy variables on the other. Such a confounding disallows any inference that task-motivated and hypnotic conditions are equivalent in essence as well as in effect.

---

[1] This paper is based on a doctoral dissertation submitted to the Department of Psychology at the University of Illinois. The author wishes to thank the thesis chairman, William Gilbert, for his continued support and guidance on this project, and John Williamsen for his untiring aid and helpful suggestions.

[2] Now at the University of Waterloo, Ontario.

Inasmuch as the influence of demand characteristics has proven to be ubiquitous for experimental subjects in general (Orne, 1962; Orne & Scheibe, 1964), the fact that control simulators will respond to them is not at all surprising. The strategy of the present experiment takes advantage of this fact and focuses on a slightly different question, namely: will hypnotic subjects respond to an alteration in demand characteristics (in the same way that simulators do) when so responding is opposed by previously administered, posthypnotic suggestions?

Through the use of posthypnotic amnesia, the present experiment will attempt to demonstrate that the effects of demand characteristics and trance variables are distinguishable. The following two working assumptions are made: (*a*) that in general, a simulator subject's failure to verbalize presumably amnesic material or to justify previous behavior is a direct effect of a group of variables subsumed under the rubric "demand characteristics," and that subsequent and appropriate manipulations of the demand characteristics can dispose such a subject to verbalize what he remembers; (*b*) that an hypnotic subject's "amnesic" and "nonjustifying" behavior is a direct effect of some variable other than demand characteristics and that subsequent manipulations of demand characteristics designed to elicit recollections or justifications from such a subject will fail in their aim. This "other" variable responsible for an inability to remember or to justify behavior is assumed to be somehow related to an altered state of the subject like that ostensibly induced by hypnosis. Let this be called the trance variable.

The above two assumptions have focused on two aspects of a subject's behavior: a failure to verbalize, which is referable to some presumably amnesic state of the organism, and the presence or absence of justifying behavior. The latter notion requires some further elaboration. For the purposes of this study, a person who justifies his previous behavior is one who claims that the behavior in question was performed voluntarily; a person who does not justify prior behavior either denies or does not admit the occurrence of the behavior, or claims that it was performed involuntarily. It seems reasonable to suppose that the presence or absence of the volitional quality of experience is related to the presence or absence of the altered state ostensibly induced by hypnosis (Austin, Perry, Sutcliffe, & Yeomens, 1963; Erickson & Erickson, 1941; Miller, Galanter, & Pribram, 1960).

The dependent variable of this experiment was subjects' verbal reports collected in an ostensibly postexperimental inquiry consisting of standardized questions about the apparent experimental (Taffel[3]) task. Use of subjects' postexperimental reports as the dependent variable has been suggested previously by Orne (1959) and Barber and Calverley (1962), and it has been utilized in a modified form by Fisher (1954) and Austin et al. (1963).

---

[3] The Taffel task (Taffel, 1955) is a task originally utilized in verbal conditioning experiments. The only equipment consists of a series of 3 × 5 index cards on each of which appears (in different order on each card) the personal pronouns I, he, she, you, we, and they. A different past tense verb is located above the pronouns on each card. The sub-

The hypotheses to be tested are as follows:

I. There will be a positive relation between the hypnotic condition and a failure to verbalize the use of "he" and "they" in a postexperimental inquiry.

II. Fewer hypnotic subjects than simulating subjects will justify their use of "he" and "they" on the Taffel task during a postexperimental inquiry.

METHOD

Subjects

The subjects were selected from a volunteer group of approximately 450 undergraduate women at the University of Illinois, most of whom were enrolled in an introductory course in psychology. Some of the subjects were enrolled in more advanced psychology courses, and a very few of them heard about the experiment from friends and volunteered their services. Subjects from this group were selected for the initial phases of the experiment on the basis of a score earned on the Harvard Group Scale of Hypnotic Susceptibility (Shor & Orne, 1962). Subjects who passed 8 items (including the posthypnotic amnesia item) out of 12, and who expressed a willingness to do further individual work in hypnosis were seen for 1 or 2 more sessions during which training and practice in attaining deep hypnosis were given. During these training sessions, no specific suggestions of posthypnotic amnesia were administered. The purpose of this deletion was to minimize later effects of subjects' expectancies concerning this phenomenon on the eventual dependent variable.

After the training sessions 10[4] out of the 12 items of the Stanford Hypnotic Susceptibility Scale Form C depth scale (Weitzenhoffer & Hilgard, 1962) were administered to each subject. A score of 7 or more was considered criterial. After both screening procedures were completed, 37 subjects remained. Eight subjects served in pilot work; 29 were run in the final study. Subjects were randomly distributed into an experimental hypnotic group (hypnotic subjects), and a control, simulating group (simulating subjects). One of the simulating subjects was highly deviant in her performance to the Taffel task, so her responses to the inquiry procedure were not analyzed. The data of another subject were lost due to a breakdown of the recording equipment.

Procedure

The general strategy of the experiment was to manipulate the experimental situation so as to elicit a "confession" from most simulating subjects concerning their voluntary use of "he" and "they," and then to see how identical procedures affected the hypnotic subjects. It was not so easy as it might seem to elicit remembering and justifying behavior from a highly susceptible, hypnotically trained subject who was consciously committed to playing the role of a posthypnotic subject. Despite the difficulties, however, a four-stage procedure was evolved that consistently resulted in the desired "confessions" by the pilot simulating subjects. This procedure was subsequently applied to the hypnotic subjects and simulating subjects in the main study as outlined below:

I. At the beginning of the final (third or fourth individual) meeting, the hypnotist (Experimenter 1) turned on the portable tape recorder which was on his desk in full view of the subject. (This was standard procedure for all of the individual practice sessions as well.) He then hypnotized the subject by means of the arm levitation technique with the suggestion that the subject would be more deeply asleep than she had ever been before by the time her hand touched her face. Experimenter 1 then read suggestions to the sleeping hypnotic subject stating in effect that (*a*) she would always begin sentences to a task, soon to be presented, with what was being called the correct response, that is, with either "they" or "he"; (*b*) that she would be unaware of this fact "just as you are unconscious of your rate of breathing or the fact your eyes blink," and (*c*) that upon awakening, she would neither remember nor be reminded of any of these suggestions. The hypnotic subject was then awakened and introduced to Experimenter 2, who presented the ostensible, experimental task. Experimenter 1 was absent from the room at this time. The simulating subjects were treated exactly the same way except that prior to the reading of the above hypnotic suggestions, they were awakened and read the following instructions:

> In a moment . . . I am going to read you exactly the same instructions hypnotized subjects are read . . . and later on I want you to pretend as though you had been hypnotized when you were read the instructions, and act just like you think the subject who had been hypnotized would act under the circumstances. Do you understand? Fine. Here are the instructions.

II. Experimenter 2, who was ignorant of the group to which any particular subject belonged, first administered standard instructions for the Taffel task. Twenty-nine instead of the usual 80 cards were presented. After the subject completed the Taffel task, Experimenter 2 presented 2 MAPS

---

ject is asked to make up sentences to the index cards, always beginning each sentence with one of the pronouns, and using the verb as the predicate.

[4] The posthypnotic amnesia item was deleted for reasons outlined in the text. The arm rigidity item was deleted because it was consistently utilized as an initial warm-up suggestion for all subjects in their individual training sessions.

pictures. The subject was asked to make up 5 sentences to each of the 2 pictures. After the subject completed the 10 sentences, Experimenter 2 left the office and Experimenter 1 re-entered.

III. Experimenter 1 sat down behind the desk and switched off the tape recorder on his desk. Simultaneously, through the use of a foot-pedal control, he switched *on* another tape recorder hidden from the subject's view in a large desk drawer. This machine thus provided verbatim records of the subsequent (Part IV) inquiry procedure. He then announced to the subject, "Well, the experiment is over. Do you have any questions about it?" No matter what the subject said, Experimenter 1 eventually worked the conversation around to the point where he could say without apparent irrelevance:

> The purpose of the experiment was to see if the way subjects make up sentences to the index [Taffel] cards has an influence on the way they make them up to the [MAPS] pictures. And the reason I had a third person administer the task is because theoretically at least I could have unwittingly biased the responses in the way I wanted them to come out.

The reason for turning off the mock tape recorder, announcing that the experiment was over, and providing the subject with a *raison d'etre* for the experiment was to alter the situation in such a way that it would no longer be perceived as experimental in nature. It was assumed that such an alteration would reduce, if not eliminate altogether, a subject's further need to perform successfully (Barber, 1962).

The subject was allowed to make any final comments, and just before she stood up, Experimenter 1, in an apparently spontaneous but actually standardized fashion, asked the subject if she could talk to another person (Experimenter 3) for a moment. This person was purportedly doing pilot research in an experiment concerning verbal behavior and had expressed interest in talking to Experimenter 1's subjects. After Experimenter 1 had made appropriate explanations to the subject and had extracted her permission to be interviewed by Experimenter 3, Experimenter 1 said, "And listen, since we're all finished here, *I want you to be completely honest with him, regardless of what I've said before.*" The purpose of the italicized portion of Experimenter 1's "closing speech" was to convince the subject that Experimenter 1 wanted her to tell everything to Experimenter 3 she could remember, *despite* any suspicions she might have had about whether the experiment was in fact over. Thus, two factors were operating to promote the subject's "confession" concerning her use of "he" and "they." First, the experiment was declared to be over, thereby removing an experimental subject's normal tendency to perform successfully from that point on. Secondly, a new and different demand was established to the effect that the subject was to be honest despite any previous instructions.

IV. Experimenter 3, on any given occasion, did not know in advance of his interview with a subject whether she was an hypnotic subject or a simulating subject. Experimenter 3 was introduced to the subject, thanked her for giving him a few extra moments, and acknowledged his good fortune in being able to use Experimenter 1's subjects as part of his pilot data. In a conversational manner, Experimenter 3 then asked the subject a series of standardized questions embedded in a matrix of irrelevant questions and comments designed to minimize the subject's suspicions. There was no attempt to adhere to a specific schedule, and the conversation proceeded in what appeared to be a spontaneous fashion. Experimenter 3 did not discourage subjects from spontaneously elaborating answers or asking questions. When Experimenter 3 did ask the pertinent questions, however, he asked them in order and in their standarized form insofar as the circumstances allowed. These interview sessions were tape recorded, and the kind of responses that the subjects gave, judged in terms of the criteria presented below, served as the dependent variable of this study. The standardized questions follow:

> (*a*) Tell me, how did you go about making up sentences to the pictures? (pause) For example, was there any regularity or consistency about the kind of sentences you made up to either of the two pictures?
>
> (*b*) Was there any similarity in the way you made up sentences to the pictures and the way you made them up to words on the index cards?
>
> (*c*) Was there any regularity or consistency in the way you made up sentences to the index cards?
>
> (*d*) From the start to the finish of each sentence, tell me step by step how you went about making up sentences to the words on the index cards.
>
> (*e*) [If there has not yet been any acknowledgment of the use of "he" and "they":] Well, one of the things you haven't mentioned, perhaps because you didn't think it was important, is the fact that you started your sentences with "he" and "they," (pause) Hmm? (pause) Do you recall now? [Regardless of answer:] I'm wondering why you didn't mention the use of "he" and "they." Several possibilities occur to me. One is that you couldn't remember the words, another is that you could remember the words but you didn't think you were supposed to mention them, and still a third is that you just didn't think it was important. Which of these three alternatives sounds the closest to the way you felt?

The questions up to this point were concerned with eliciting information about posthypnotic amnesia. In order to be counted as amnesic, a subject had to say nothing about the use of "he" and "they" through Question *d, and* in response to Question *e*, say that she could not remember using the words.

The following questions were concerned with the ability of a subject to justify her behavior and were asked only if the subject immediately acknowledged or ultimately recognized her prior use of "he" and "they."

> (*f*) *At the time* you were actually doing the task did you realize that you were starting all your sentences with "he" and "they"? [If "yes," Experimenter 3 went on to Question *g*; if "no," then f_1 was asked.]
>
> (f_1) When you say "no" to this question, it's possible to mean one of two things. One possibility is that even though you knew you were using "he" and "they" it was really very natural and easy to do this, and therefore you simply didn't have to think much about it as you went along. The second possibility is that you really didn't have the slightest idea that you were starting your sentences with "he" and "they." If this were the case, you were probably very surprised to discover that this was what you had done. Which of these possibilities seems closer to the truth, the first one or the second one? [If the subject replied with the former, it counted the same as if she had said "yes" to question *f*. If the subject replied with the latter, no further questions were asked.]
>
> (*g*) If you were aware of using "he" and "they," did you realize *at the time* why you were using the words?
>
> (*h*) *At the time* did you use the words involuntarily or voluntarily? And by voluntarily I mean did you use them on purpose because you knew you were supposed to, that is, because they had been designated as correct for this experiment?

The subjects were said to justify their behavior only if they claimed that their use of "he" and "they" was voluntary in response to question *h*. The subjects were said not to justify their behavior who replied to question *h* that they had responded involuntarily, or who were never asked this question because of its irrelevancy in light of previous answers. It was *assumed* that only simulating subjects would justify their behavior and the validity of this assumption was tested by seeing whether hypnotic subjects and simulating subjects could be accurately identified on the basis of this criterion.

After Experimenter 3 completed his interview, he left the office and Experimenter 1 returned. If the subject was a simulating subject, Experimenter 1 thanked her for helping Experimenter 3, and admonished her to say nothing about what had happened during the final session. If the subject was an hypnotic subject, Experimenter 1 rehypnotized her, removed the suggestion to begin sentences with "he" and "they," and awakened her. Neither hypnotic subjects nor simulating subjects were ever told the actual purpose of the experiment at this time. This precaution was taken to minimize the importance of any information leaks that might occur despite Experimenter 1's plea for silence.

RESULTS

There was no significant difference in the hypnotic subjects' and the simulating subjects' tendency to use "they" and "he" on the Taffel task nor was there any significant difference in the frequency with which hypnotic subjects and simulating subjects used the "correct" pronouns to begin sentences to the MAPS pictures.

Although the effect of hypnosis did not produce significant differences on a task explicitly defined as experimental, there were quite striking differences in the "postexperimental" reports of hypnotic subjects and simulating subjects. Before reporting the actual findings, however, a word should be interjected about how the subjects' reports were classified. It might seem that the explicitness of the criteria for amnesia and justification would disallow any equivocation in judging their occurrence for any given subject. Indeed, this was the case for the vast majority of the subjects. However, a small minority of subjects responded in such a way that made it difficult to ascertain unequivocally whether or not they fell into one response category or another. This occasional ambiguity occurred for both the dimensions of amnesia and justification. In order to arbitrate this matter, two independent judges assessed all of the tape-recorded responses. The data of the judge least favorable to the experimental hypotheses are reported below. In confirmation of Hypothesis I, there was a high relationship ($\phi = .62$; $p < .0005$) between the hypnotic condition and the lack of verbalization of "he" and "they" during the "postexperimental" inquiry. All of the 13 simulating subjects verbalized the use of "he" and "they" at this time whereas 8 out of 14 hypnotic subjects did not. Subjects' use of "he" and "they" was agreed upon 89% of the time by the two judges. In confirmation of Hypothesis II, the use of justification as a criterion in discriminating hypnotic subjects succeeded 25 out of 27 times. All 13 simulating subjects justified their behavior; 12 out of 14 hypnotic subjects did not (one-tailed $\chi^2 = 20.05$; $p < .00005$). Two independent raters agreed in their classification of justifiers versus nonjustifiers 92% of the time.

Three possible objections might be raised about the present findings. First, it might be argued that Experimenter 1 was *not* naive about the condition of the subject and that this knowledge could have subtly and selectively affected the way he vocalized his closing speech (Rosenthal, 1964). In an attempt to circumvent this criticism entirely, Experimenter 1 did not give the closing speech during early pilot work. Instead, Experimenter 3 announced the end of the experiment after ostensibly collecting data for Experimenter 1. Following this announcement, Experimenter 3 plied the subject with some "informal" questions of his own, the answers to which were utilized in making judgments concerning amnesia and justification. Now, the research program consisted of seeking and finding a procedure that would consistently elicit "remembering" and "justifying" behavior from simulating subjects and then to see how the identical procedures affected the hypnotic subjects. Unfortunately, under these initial pilot conditions, simulating subjects did not "confess" even though they were convinced the experiment was over. Their rationale for continued withholding was a belief that only the hypnotist could release really hypnotized subjects from previous suggestions by announcing the end of the experiment. The effect of these confounding expectancies obviously had to be eliminated for *both* hypnotic subjects and simulating subjects, so Experimenter 1 thenceforth announced the experiment's finale.

In lieu of obviating completely the possibility that Experimenter 1 would convey differential treatment effects through subtle and confounding vocal cues, it was possible to test for their existence. The 27 repetitions of Experimenter 1's closing speech were tape-recorded and four independent judges[5] were asked to judge from these recordings whether Experimenter 1 was talking to an hypnotic subject or a simulating subject on any given occasion. Three judges met with totally chance success in their judgments, but one judge did discriminate successfully ($p < .05$). When this judge was asked what his rationale for categorizing subjects was, he replied that Experimenter 1 seemed more convincing and persuasive with the hypnotic subjects! If this were in fact true, that is, if his success were not simply improbable, it seems plausible that hypnotic subjects would have been more prone to "confess" than simulating subjects all other things being equal. In any event, it seems unreasonable to suppose that naive experimental subjects could have been consistently influenced by vocal cues the very presence of which were so unreliably assessed by sophisticated judges set to discriminate.

A second possible objection to the present findings concerns the possibility that the inquiry procedure might have aroused subjects' suspicions that the experiment was not over, and that this suspicion might in some way be used to explain the present results better than the independent variable. To check out this possibility, Experimenter 1 telephoned each subject after all the subjects had been seen for the last time. He asked, "By the time Experimenter 3 was actually asking you his questions, did you have any notion that in fact my experiment was not over?" Nine hypnotic subjects and six simulating subjects said "No," three hypnotic subjects and five simulating subjects said "Yes."[6] (Fisher exact probability test insignificant at .05 level). At least by their own account, the hypnotic subjects and simulating subjects did not differ significantly in their perception that Experimenter 3's questions constituted a *post*experimental inquiry. It follows that insignificant differences in how the inquiry procedure was perceived could not account for the highly significant differences of hypnotic subjects' and simulating subjects' verbal reports to the inquiry.

A third objection that might possibly be levelled at the present experiment is this. Actually, there were two differences in the way the hypnotic subjects and the simulating subjects were treated: hypnotic subjects but not the simulating subjects received their suggestions while hypnotized; simulating subjects but not hypnotic subjects received special instructions to behave later *as if* they had been hypnotized when they received the subsequent suggestions. Since there were two differences in the way the groups were treated, any significant results could as easily be attributed to instructional differences as to hypnosis.

---

[5] One senior psychologist, two advanced graduate students in clinical psychology, and one PhD in electrical engineering served as judges.

[6] Two hypnotic subjects and two simulating subjects were not included in this computation because their answers disallowed objective classification.

This is a logical argument, but not an entirely reasonable one when the purpose of the simulating instructions is clarified. If these instructions had been deleted, then some of the simulating subjects would have been apt to make wrong assumptions about how they were to behave. For example, it would be quite plausible for a simulating subject, unprepared by simulating instructions, to assume that subsequent hypnotic suggestions were not supposed to "take" simply because both he and the hypnotist knew he was not hypnotized upon receipt of the suggestions. Such an assumption by a simulating subject could be expected to lead to obviously nonhypnotic behavior. The explicitness of simulating instructions thus served to maximize the likelihood of *all* simulating subjects behaving hypnotically insofar as this behavior is determined by the experimental demand characteristics. The purpose of maximizing this source of hypnotic behavior in simulating subjects was to minimize the differences between the two groups insofar as this difference might be a function of the hypnotic subjects' sensitivity to the demand characteristics (Fisher, 1954). In short, the function of the simulating instructions was to enhance the likelihood of accepting the null hypothesis if in fact hypnotic behavior were nothing but responsiveness to demand characteristics. Therefore, any significant differences between groups that did appear seem to be the result of something other than simulating instructions, namely, the presence of the trance variable in hypnotic subjects only.

It must be admitted that the above line of thinking is only reasonable, and that the merely reasonable can occasionally be as misleading as the merely logical. For example, it might be true that demand characteristics for hypnotic subjects were totally different from those of the simulating subjects simply because the hypnotic subjects knew or believed themselves to be hypnotized and the simulating subjects knew that they were not. This is a somewhat vexing possibility not easily accessible to experimental control. Barber and Calverley (1964b) suggest that a slight variation of this possibility is the case, but their conclusions in this and similar experiments are suspect due to faulty methods described below. Anecdotal evidence suggests, however, that such belief or knowledge is not a crucial demand variable in hypnosis. Occasionally, one hypnotizes a highly susceptible subject who afterwards claims (oftentimes with some vehemence) that he knew at the time and knows now that he in fact was not hypnotized. This claim typically does not square with the clinical facts, for example, unacknowledged and spontaneous posthypnotic amnesia. Evidently in these cases it is not the demand characteristics supplied by the knowledge or belief of being hypnotized that eventuates in hypnotic behavior.

DISCUSSION

Inasmuch as differences due to experimental demand characteristics were minimized, the results of the present investigation seem to constitute one of the more convincing experimental demonstrations of the "altered state" theory of hypnosis. The success in this regard was made possible in part by avoiding a methodological pitfall common to investigations which have concluded or implied that the concept of an hypnotic state is expendable. The flaw that has prompted this conclusion is simply stated as follows: typically, the experimenter provided too much information in the initial (hypnotic, simulating, task motivating, etc.) suggestions about what kind of (dependent variable) behavior was expected. If, for example, both hypnotized subjects and task-motivated (simulating) control subjects are told they are to see an (hallucinatory) cat on their laps or hear *Jingle Bells* played on an imaginary phonograph (Barber & Calverley, 1964a) then both groups of subjects have a good idea of what is expected of them: they are supposed to say that they saw a cat and heard the song. Given the admission of hallucinations by both hypnotic and task-motivated subjects, it is impossible to know if the hypnotic subjects are *simply* responding to the demand characteristics as the control subjects are, or if they so respond because they "see" and "hear" the suggested phenomenon as a function of having received these suggestions in a hypnotic trance.

Barber and his co-workers are particularly vulnerable to this criticism. He has presented data from many experiments all of which lead him to conclude that simply giving direct suggestions to waking, task-motivated subjects is sufficient to elicit behavior that cannot be differentiated from "hypnotic" behavior. If this were all he concluded, there would be little to disagree with. However, he further implies that since hypnotic subjects do not *behave* differently from task-motivated controls, the *basis* of their behavior must be similar. This does not seem to be a legitimate conclusion.

It is very important to be clear about this. A person can voluntarily perform a tic that to all appearances is indistinguishable from an involuntary, nervous tic. The behavior is the same; the basis of the behavior quite obviously is not. The actual differences underlying these two kinds of tic behavior are of paramount theoretical and practical importance, for example, the unconscious as opposed to conscious origins of the nervous tic, the implications this has for treatment, and so on. Equally important issues are at stake in the experimental study of hypnosis. Consequently, insofar as one is grappling with questions about the nature of hypnosis, there is a point of diminishing returns implied by the exclusive use of paradigm that tends, by its very nature, to emphasize the similarities between real and simulated (or task-motivated) hypnotic behavior rather than to clarify the nature of the differences underlying them. That hypnotic and control subjects *do* behave similarly is certainly a valuable finding, particularly when it is not an expectable result on strictly *a priori* grounds. The question remains, however, whether or not the similarity in behavior reflects a similarity in the *basis* of behavior.

There seem to be at least three competing answers to the question of why hypnotic subjects and their controls do in fact behave similarly: (*a*) the control subjects are somehow hypnotized by virtue of having received their suggestions in a motivated state (a possibility that seems almost too remote to consider); (*b*) hypnotic subjects are merely task motivated and/or responsive to the experimental demand characteristics (a possibility that Barber seems to favor); (*c*) control subjects simply respond in a more or less purposeful manner to the task demands, whereas the hypnotic subjects' behavior is determined both by demand variables *and* by state or trance variables.[7] The present investigation seems rather unequivocally to confirm this last alternative. Previous investigations, however, utilizing the usual hypnotic-simulator paradigm have not really legitimized a choice among these options. Their strategy has generally consisted solely of proving the null hypothesis that hypnotic and simulator subjects can behave similarly. But no matter how many times this same paradigm is used to "prove" the null hypothesis of behavior equivalence, it will never suffice to accept or reject the null hypothesis concerning the underlying bases of similar behavior. To maintain contrariwise, that is, to conclude that because behavioral manifestations of hypnotic and simulating subjects are similar, the behavioral substrate must be, is a little like saying that the person who voluntarily performs a tic must be as neurotic as the genuine tiqueur.

The present investigation suggests that earlier studies showing no difference between hypnotic subjects and their controls did so for a very good reason: they compared groups in terms of behavior that was entirely explicable on the basis of demand characteristics alone, thereby minimizing the possibility of isolating any special effects the hypnotic state may have had. Whether hypnotic subjects behaved the way they did simply as a response to demand variables or whether for uniquely hypnotic reasons could not, therefore, be determined.

---

[7] These alternatives are not necessarily exhaustive. Indeed, much of the literature on social persuasion and conformity behavior seems pertinent to Barber's methods. Oftentimes, for example, Barber presents what amounts to false norms to his task-motivated controls, telling them, for instance, that so far *everyone* who has really tried to see a cat on their lap has been able to do so. A subject's subsequent admission that he indeed did see a cat bears a certain resemblance to the behavior of subjects in an Asch type experiment whose responses on a variety of judgment tasks are altered greatly by the influence of false norms (Crutchfield, 1955). It has been shown by factor analysis that this general kind of responsiveness to social norms bears little or no relation to hypnotic susceptibility (Moore, 1964).

The present study obviated this interpretive dilemma by making critical alterations in the hypnotic-simulator paradigm usually employed. In previous studies, the influence of trance and task demand variables has been in the same direction. This merger confounded their effects in a way that obscured any influence the trance variable may have possessed. In this study, therefore, the directions of influence of trance variables and subsequent demand variables were purposely antagonized, thereby allowing a legitimate inference from the dependent variable behavior concerning which of the variables was prepotent.

The following is a summary account of the revamped hypnotic-simulator paradigm as it was employed in this experiment. Neither the hypnotic subjects nor the simulating subjects received any specific information about how to behave after the announced conclusion of the experiment. Thus, demand characteristics inherent in the initial suggestions, and specific to the dependent variable behavior were minimized. The subsequent termination of the experiment, together with a directive to be honest despite previous suggestions altered the situation for the simulating subjects so that it became all right for them to "confess." The fact that an identical alteration in the demand characteristics did not similarly affect the hypnotic subjects is, of course, the critical finding of this investigation. The effects of receiving suggestions under hypnosis evidently take precedence over subsequent, countermanding alterations in the demand characteristics. It is reasonable to conclude that hypnotic behavior is not wholly reducible to acting in accordance with demand characteristics, and that hypnosis seems in part to be an altered state within which suggestions have a peculiarly potent effect.

REFERENCES

ÅS, A. Hypnotizability as a function of nonhypnotic experiences. *Journal of Abnormal and Social Psychology,* 1963, **66,** 142–150.

AUSTIN, MARGARET, PERRY, C., SUTCLIFFE, J. P., & YEOMENS, N. Can somnambulists successfully simulate hypnotic behavior without becoming entranced? *International Journal of Clinical and Experimental Hypnosis,* 1963, **11,** 175–186.

BARBER, T. X. Towards a theory of hypnosis: Posthypnotic behavior. *Archives of General Psychiatry,* 1962, **7,** 321–342.

BARBER, T. X., & CALVERLEY, D. S. "Hypnotic behavior" as a function of task motivation. *Journal of Psychology,* 1962, **54,** 363–389.

BARBER, T. X., & CALVERLEY, D. S. An experimental study of "hypnotic" (auditory and visual) hallucinations. *Journal of Abnormal and Social Psychology,* 1964, **68,** 13–20. (a)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of defining the situation as hypnosis and defining response to suggestions as easy. *Journal of Abnormal and Social Psychology,* 1964, **68,** 585–592. (b)

BARBER, T. X., & HAHN, K. W. Physiological and subjective responses to pain producing stimulation under hypnotically-suggested and waking-imagined "analgesia." *Journal of Abnormal and Social Psychology,* 1962, **65,** 411–418.

CRUTCHFIELD, R. S. Conformity and character. *American Psychologist,* 1955, **10,** 191–198.

ERICKSON, M. H. An experimental investigation of the hypnotic subject's apparent ability to become unaware of stimuli. *Journal of General Psychology,* 1944, **31,** 191–212.

ERICKSON, M. H. Deep hypnosis and its induction. In Le Cron (Ed.), *Experimental hypnosis.* New York: Macmillan, 1952.

ERICKSON, M. G., & ERICKSON, E. M. Concerning the nature and character of post-hypnotic behavior. *Journal of General Psychology,* 1941, **24,** 95–133.

FISHER, S. The role of expectancy in the performance of post-hypnotic behavior. *Journal of Abnormal and Social Psychology,* 1954, **49,** 503–507.

LONDON, P., & FUHRER, M. Hypnosis, motivation, and performance. *Journal of Personality,* 1961, **29,** 321–333.

MARCUSE, F. L. *Hypnosis: Fact and fiction.* Baltimore: Pelican, 1959.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

MOORE, ROSEMARIE K. Susceptibility to hypnosis and susceptibility to social influence. *Journal of Abnormal and Social Psychology,* 1964, **68,** 282–294.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology,* 1959, **58,** 277–299.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

ORNE, M. T., & SCHEIBE, K. E. The contribution of nondeprivation factors in the production of sensory deprivation effects: The psychology of the "panic button." *Journal of Abnormal and Social Psychology*, 1964, **68**, 3–12.

ROSENTHAL, R. Experimenter outcome-orientation and the results of the psychological experiment. *Psychological Bulletin*, 1964, **61**, 405–412.

SHOR, R. The frequency of naturally occurring "hypnotic-like" experiences in the normal college population. *International Journal of Clinical and Experimental Hypnosis*, 1960, **8**, 151–163.

SHOR, R. E., & ORNE, E. C. *Harvard Group Scale of Hypnotic Susceptibility*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.

SIDIS, B. Are there hypnotic hallucinations? *Psychological Review*, 1906, **13**, 239–259.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology*, 1955, **51**, 496–501.

WHITE, R. W. A preface to the theory of hypnotism. *Journal of Abnormal and Social Psychology*, 1941, **36**, 477–505.

WEITZENHOFFER, A. M. *Hypnotism: An objective study in suggestibility*. New York: Wiley, 1953.

WEITZENHOFFER, A. M. *General techniques of hypnotism*. New York: Grune & Stratton, 1957.

WEITZENHOFFER, A. M. The nature of hypnosis, Part II. *American Journal of Clinical Hypnosis*, 1963, **6**, 40–72.

WEITZENHOFFER, A. M., & HILGARD, E. H. *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto: Consulting Psychologists Press, 1962.

(Received February 15, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 52–59

INDIVIDUAL DIFFERENCES IN DREAM RECALL [1]

HELEN B. LEWIS, DONALD R. GOODENOUGH, ARTHUR SHAPIRO, AND IRVING SLESER [2]

State University of New York, Downstate Medical Center

Ss who typically fail to remember dreams at home (nonreporters) and Ss who frequently remember dreams (reporters) slept in the laboratory for 4 nights each. Gradual or abrupt awakenings were made at each EEG Stage-1 REM (dream) period. Although nonreporters and reporters did not differ in REM-period frequency or EEG patterns during sleep, nonreporters did report dreaming less frequently following REM-period awakenings. Ss showed self-consistency in frequency of dream reporting and in type of failure to report. Some nonreporters typically failed to remember any content; others typically said they were awake and thinking. Comparisons among reporters and subgroups of nonreporters for eye-movement frequency, arousal threshold, and dreamlike-report content indicate that it may be useful to distinguish different kinds of nonreporters.

This study is one of a series of studies comparing subjects (*S*s) who typically remember dreams with *S*s who claim that they rarely dream, in an effort to understand the basis for these individual differences in dream recall. These studies make use of the Aserinsky-Dement-Kleitman (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957a, 1957b) EEG and eye-movement criteria for determining when dreams occur. They are based on the working assumption that awakening a *S* during a Stage-1 EEG rapid eye-movement period (REM period) is awakening him while he is in a dream state.

The first study in the series (Goodenough, Shapiro, Holden, & Steinschriber, 1959) was designed to test the accuracy of the claim made by some people that they rarely or never dream. A group of these *S*s (nonreporters) and a group of *S*s who claimed that they recalled dreams almost every night (dream reporters) were studied in the laboratory during their normal sleep period. The two groups did not differ significantly in the frequency of rapid eye-movement (REM) periods. When awakened from REM periods, every *S*, including the nonreporters, related at least one dream in some detail. However, the nonreporters related significantly fewer dreams than the reporters, even under these laboratory conditions.

In studies of uninterrupted sleep, Antrobus (1962) and Antrobus, Dement, and Fisher (1964) have also found approximately the same number of REM periods in reporters and nonreporters. However, they did find that REM periods tended to be somewhat shorter among the nonreporters. This difference in REM length is of great interest but does not seem large enough in itself to account for individual differences in dream reporting.

If we assume that dreams occur during REM periods, these findings seem to indicate that so-called "nondreamers" dream about as often as other *S*s, but for some reason, frequently fail to report dreams even when questioned immediately after the REM period is interrupted.

In an attempt to understand these failures to report a dream on awakening from a REM period, the *S*s' own descriptions of what they experienced were examined. An impressive variety of phenomenologically different types of experiences were described by the *S*s. In some cases, a *S* insisted that he had experienced something but could not recall the content of his experience. In these cases, he often claimed that the experience was lost—got "sucked away" somehow—when the bell rang. In other cases, the *S* reported that his "mind was a blank." On still other occasions, the *S* reported that he was "thinking" rather than dreaming, or that he had been awake rather than asleep while some cognitive experience was occurring.

---

[1] This study was conducted in the Psychology Laboratory and the Psychiatric Treatment Research Center of the Department of Psychiatry, and was supported by the National Institutes of Health, United States Public Health Service, grant nos. MH-03885, MH-05518, and MH-K3-16,619.

[2] Now at the Research Center for Mental Health, New York University.

The occurrence of such variety raises the possibility that the reason why the *S* failed to report a dream might be different in different cases. For example, when an *S* said that he was awake and thinking rather than asleep and dreaming, he actually might have been in a relatively light, hypnagogic-like sleep state during the REM period. On the other hand, when a *S* was unable to report any content, a recall failure might have been involved. If these types of reports do have different bases, one might expect them to occur under different conditions and, perhaps, in different *S*s.

Some evidence for the view that the different types of reports occur under different conditions was found in previous studies (Goodenough, Lewis, Shapiro, Jaret, & Sleser, 1965a; Goodenough, Lewis, Shapiro, & Sleser, 1965b; Shapiro, Goodenough, & Gryler, 1963; Shapiro, Goodenough, Lewis, & Sleser, 1965). For example, there is some evidence that many thinking reports may be descriptions of arousal artifacts in situations where the *S* fails to recall the dream which presumably occurred during the preceding REM period. The present study was designed to check on the possibility that different types of failure to report a dream might occur in different *S*s, and that different *S*s might be nonreporters for entirely different reasons. Specifically, it was hypothesized that some nonreporters typically fail to report any content and others typically report "thoughts" rather than dreams when REM periods are experimentally interrupted. In addition, it was possible to explore a number of hypotheses about the nature of these types of failure to report dreams by studying some correlates of individual differences in dream reporting.

One possibility considered was that dream recall might be more difficult for *S*s who sleep relatively deeply during their REM periods. In previous studies (Goodenough et al., 1965b; Shapiro, Goodenough, Biederman, & Sleser, 1964), it has been found that reports in which the *S* believes he was asleep but fails to recall any content tend to come from REM periods early at night during which the *S* is evidently more deeply asleep than is usual for him. In the present study, it was hypothesized that *S*s who frequently fail to report any content would be typically more difficult to awaken from their REM periods than *S*s who usually recall their dreams.

Several alternative possibilities were considered in an attempt to account for the thinking reports. It seemed possible that some *S*s might have what other people call dreams but which they, for some reason, label as thoughts. This hypothesis was tested by comparing the narratives of the reports given by *S*s who typically label their experiences as thoughts with the narratives given by *S*s who typically label their experiences as dreams on a number of characteristics (e.g., bizarreness, visual imagery, etc.).

The possibility was also considered that REM-period experiences are actually more thoughtlike in nature in some *S*s. It seemed reasonable to expect that borderline states between waking and sleeping might involve hypnagogic-like cognitive activity which the *S* would label as thoughts rather than dreams. One of the hypotheses tested was that thinking reports are given by *S*s whose REM periods occur in a particularly light sleep, as judged by arousal thresholds and alpha activity in the EEG records. A second hypothesis, based on the assumption that conjugate eye movements reflect visual scanning during the REM-period experience (Dement & Kleitman, 1957a), was that *S*s who typically give thinking reports have REM periods during which eye movements are relatively infrequent.

METHOD

The data to be reported are based on 46 young adult male *S*s who were paid for their participation in the study. All but 7 of these were students who volunteered to sleep in the laboratory for studies on the effects of the method of awakening on dream reporting (Goodenough et al., 1965a; Shapiro et al., 1963; Shapiro et al., 1965). All of the *S*s were selected on the basis of their response to a questionnaire designed to elicit information on how often they recalled their dreams. Sixteen *S*s who claimed that they recalled their dreams almost every night were tentatively classified as dream reporters, and 30 *S*s who claimed that they recalled dreams no more often than once a month were tentatively classified as nonreporters.

The *Ss* were studied on at least four nonconsecutive nights in the laboratory during the hours when they usually slept. In a few cases, a fifth night was added when data from the first 4 nights were sparse. The four (or five) experimental nights were generally run at the rate of one a week.[3]

During the interval between laboratory visits (or, in a few cases, after completion of the visits), the *Ss* were asked to keep a diary of their home dreams. Home-report forms covering from 1 to 5 weeks were received from all but one *S*. These home reports were used to supplement the questionnaire data in evaluating the frequency of home-dream recall. When grossly conflicting results were obtained from the questionnaires and dream diaries, the *Ss* were excluded from all dream reporter/nonreporter comparisons. For these comparisons, the final groups included 12 dream reporters who recalled dreams at home more often than once every third night by both criteria, and 26 nonreporters who recalled dreams at home less often than once every third night by both criteria.[4]

The equipment and procedures used have been described in detail elsewhere (Goodenough et al., 1959; Shapiro et al., 1965).

Before each laboratory run, each S was instructed, recording electrodes were attached, and a 90-second EEG (fronto-occipital) record was taken before the *S* went to bed. The *S* was told that a bell would ring several times during the night, sometimes abruptly and sometimes with gradually increasing intensity. He was asked to pick up a bedside telephone (which automatically shut off the bell) and to describe any experience he had before the bell rang. He was told that *E* would question him over an intercom, but that he was to report his experience in as much detail as possible before *E* said anything.

During the course of sleep, continuous EEG and horizontal eye-movement potential recordings (EOG) were taken. In order to reduce variability of the length of REM period preceding each awakening, the experimental interruption was made between 5 and 10 minutes after the onset of each REM period. Half of the interruptions were abrupt and half were gradual for each *S*.

For the gradual awakenings, the initial bell stimulus was generally presented at about 20 *db* (ref. .0002 dynes/cm^2) and increased in 6-*db* steps until the *S* responded. At each step, the stimulus lasted for approximately 10 seconds, with a 5-second interval between steps. The intensity reached by the bell, before the *S* responded by picking up the phone, was used to define the arousal threshold for the REM period. For the abrupt awakenings, the stimulus was presented at about 80-*db* intensity. After the final awakening in the morning, a second waking-EEG record was taken.

In order to obtain a quantitative measure of REM-period eye-movement activity, each 3-second interval of the REM period was scored for the presence or absence of rapid, conjugate eye movements, and the percentage of 3-second intervals containing at least one movement was calculated for each period. The mean of these scores was then computed to yield a measure of the average activity for each *S*. Each 10-second interval during a REM period was also scored for the presence or absence of clearly defined alpha activity, and the mean percentage of these intervals for all REM periods observed in an *S* was used to describe his REM-period EEG. Percentage time alpha was also computed for the standard samples of waking EEG record.

The transcripts of the *Ss'* reports were first divided into two classes: those in which the *S* reported some cognitive content, and those in which no content was recalled. The *content* reports were then subdivided into those in which the *S* said that he was *dreaming*, and those in which he said that he was *thinking*. In all but a few cases, the *Ss* said that they were awake or in a semi-sleep state when they gave thinking reports. When *no content* of a cognitive nature was recalled, the reports were subdivided, depending on whether the *S* said that he was *awake* or *asleep* when the bell rang.

Content analyses of the dream and thought narratives were also conducted. It was expected that *Ss* who typically give thinking reports, in contrast with dreaming reports, would more often describe narratives which refer directly to the laboratory situation, would not involve the dreamer in the experience, would be less bizarre, have less visual imagery, and less motor activity in the imagery. In addition, word counts were conducted on the narratives, with the hypothesis that *Ss* who give thinking reports would describe experiences that are shorter, that is, less detailed. The details of the scoring procedure for these characteristics of dream content are described elsewhere (Goodenough et al., 1965b).

The content analyses were conducted by a judge who had not seen the *Ss* and had no other information about them. In the transcripts she was given, all identifying information was blocked out, and the *S*'s comments concerning the thoughtlike or dreamlike quality of the experience and sleep depth were excluded. In addition, parts of the transcript which included obvious incorporations of the awakening stimulus were censored.

---

[3] After the laboratory runs were completed, most of the *Ss* were recalled for daytime testing. The tests included the Draw-A-Person test, the Rorschach, and three perceptual tests (the Rod-and-Frame Test, the body-adjustment series of the Tilting-Room Tilting-Chair Test, and the Embedded-Figures Test) taken from Witkin's battery of field-dependence measures (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). Each *S* also received an interview, designed to elicit material that might be useful for personality evaluation. These data were collected as part of a later study of the personality correlates of dream recall and are not reported here.

[4] The mean percentage of home nights with at least one dream recalled was 52 (about once every other night) for the *Ss* who claimed on the questionnaire that they dreamed almost every night, and 19 for *Ss* who claimed that they dreamed no more often than once a month ($p < .01$).

RESULTS AND DISCUSSION

The course of sleep in terms of EEG stages and the distribution of eye-movement periods during the night followed the pattern described by Dement and Kleitman (1957a). As in our earlier study (Goodenough et al., 1959), the nonreporters had REM periods about as frequently as did the dream reporters. The mean number of REM periods per night was 4.2 for the dream-reporter group, and 4.0 for the nonreporter group. (Mean time asleep per night was approximately 400 minutes for both groups.) In our earlier study, we also found that nonreporters had more REM-period alpha activity (defined as the percentage of all 20-second intervals during REM periods which contained alpha bursts) than did dream reporters. However, in the present study, we failed to replicate this observation. The only variable considered in this study which significantly related to the amount of REM-period alpha was the percentage of time alpha for the EEG records taken under the standard waking conditions. *S*s who had prominent alpha activity when awake and relaxed with eyes closed also tended to have relatively prominent alpha during their REM periods ($r = +.62$; $N = 46$; $p < .01$).

No evidence could be found, then, that dream reporters and nonreporters differed in REM-period frequency or EEG patterns during sleep.[5]

Dream Reporting From REM-Period Awakenings

As has been found in many previous studies, the *S*s reported some dream (or thought) content from the majority of REM-period awakenings. Each of the nonreporters gave at least one report with some cognitive content. As might be expected, however, nonreporters recalled cognitive content less often than reporters under these conditions. For reporters, the percentage of awakenings which yielded some content was 82; for nonreporters, the percentage was 66 ($p < .05$). Even when the nonreporters recalled some content, they labeled their experience a thought significantly more often than did reporters. A mean of 29% of the content reports were labeled thoughts by the nonreporter group, but only 7% were so labeled by the reporters ($p < .05$). Table 1 summarizes these data.[6]

TABLE 1

DISTRIBUTION OF "THINKING" AND "DREAMING" REPORTS FROM REM-PERIOD AWAKENINGS IN REPORTERS ($N = 12$) AND NONREPORTERS ($N = 26$)

Subject classification	Mean percentage of reports with content		Mean percentage of reports without content	
	Labeled as thoughts	Labeled as dreams	Labeled as awake	Labeled as asleep
Reporters	07	93	19	81
Nonreporters	29	71	31	69
Difference	−22*	22*	−12	12

* Significance of difference, $p < .05$.

These results agree with previous findings that nonreporters, in general, report dreaming less often following REM-period awakenings. If we assume that dreams occur during all REM periods, then these studies suggest that the essential difference between dream reporters and nonreporters must be in the reporting of dreams rather than in the frequency of dreaming.

Individual Differences in the Type of Report

For each of the 46 *S*s, the percentage of REM-period awakenings yielding reports of each type was computed for his first and second nights in the laboratory and correlated with comparable percentages computed for

---

[5] No meaningful measure of REM-period duration could be obtained, since REM periods were experimentally terminated after 5 to 10 minutes, in accordance with the design of the study.

[6] The percentages given in Table 1 were computed for each *S* separately, and the mean of the percentages is used to describe the results. For the entries in the columns headed *Mean % of reports with content and Mean % of reports without content,* the number of "with content" or "without content" awakenings was used as the denominator. In general, for all data analyses given in this paper, the mean value for all observations on an *S* was computed first, and the group means are means of these subject means. All *N*s given are *N*s of *S*s.

TABLE 2

RELIABILITY COEFFICIENTS (NIGHTS 1 & 2 VERSUS 3 & 4) FOR EACH REPORT TYPE

Report type	Correlation
Total no-content reports	+.78
Content; dreaming	+.77
Content; thinking	+.59
No content; asleep	+.80
No content; awake	+.56

the third and fourth nights in the laboratory. Reliability coefficients (corrected by the Spearman-Brown formula) are presented in Table 2.

All of the correlations were significant ($p <$ .01). Despite the fact that the number of awakenings on each *S* was relatively small, the data indicate some degree of consistency in response for a given person, and marked individual differences in the characteristic reports given from REM-period awakenings in the laboratory. This finding confirms one of our starting hypotheses, namely, that there would be individual consistency in the type of report given.

There were four nonreporters who gave no-content asleep reports at least half the time, and five nonreporters who recalled some content but labeled their experience a thought at least half the time. All of the dream reporters and 10 of the nonreporters recalled dreams after at least half of their REM-period awakenings. For the seven remaining nonreporters, the REM-period reports were more evenly distributed among the different types.

The finding that some *S*s often fail to report any content from the REM-period awakenings but almost never say they were thinking, while other *S*s almost always give some detailed content but typically say they were thinking rather than dreaming, contributes to the view that the different types of failure to report a dream experience may have a qualitatively different basis.

Relationships Between Dream Reporting and the Content of the Narratives

In previous studies (Goodenough et al., 1965a; Goodenough et al., 1965b), we have found for most *S*s that when an occasional thinking report is given, it is usually less bizarre, less active, less often involves visual imagery, less often includes the self as a protagonist of the action, and tends to be shorter than dream reports obtained from the same *S*. On the basis of the content of the narrative, an external judge usually agreed with the *S* that he was sometimes thinking and other times dreaming.

This agreement did not exist for the five nonreporters who typically gave thinking reports. These *S*s appeared willing to label bizarre experiences as thoughts. The following verbatim transcript of a report given by one of these *S*s is an extreme example of a thinking report whose narrative seems dreamlike in every way:

> *S:* I was awake and thinking of going over to a candy machine. I began to put the dime into the machine and more dimes—three more—came out in the front. So I took the dime back out and put it in again and three more dimes came to the front. I did this four or five times. Then I put the dime almost all the way in and all of a sudden this pile of pennies came flowing out—an endless stream practically. I put the pennies in my pocket. This soda machine was in the classroom. I asked the teacher if she minded and she said no. I just sat there and put the pennies in. This classroom was what you call a student adviser seminar, which they have at Brooklyn College for certain people who advise groups. It seems the machine was built over an oil well, and there were two bubbling wells there. That's all.
>
> *E:* You say you were awake and thinking this?
>
> *S:* Yes. It's very hard to tell whether you're awake or sleeping.
>
> *E:* Any more details?
>
> *S:* I didn't want to share the pennies with anybody so I asked a girl if I could borrow her hat and I was going to pour everything in. Except it seems the hat was made of some sort of rubber and I couldn't pour it in because there was no room—it came all the way up to the top of the hat inside. It was a white hat.

Since naratives of this sort appeared to be so dreamlike to an external judge, the spontaneous comments given by these *S*s in describing their experiences were particularly interesting. Often these *S*s felt they were in control of the content of their experience. For example, one *S* said, "This was sort of a daydream I could have controlled. I could have started thinking things and if I wasn't satisfied, I could change it." Other *S*s reported that they maintained their orientation with respect to time and place ("I'm aware of

TABLE 3

COMPARISONS AMONG DREAM REPORTERS AND DIFFERENT TYPES OF NONREPORTERS IN TERMS OF REM-PERIOD CHARACTERISTICS

REM-period characteristic	Dream reporters ($N = 12$)	Nonreporters who typically gave		
		Thinking reports ($N = 5$)	No-content asleep reports ($N = 4$)	Dream reports ($N = 10$)
Mean threshold (*db*)[a]	45	32	45	63*
Mean eye-movement percentage	22	15*	21	22

[a] The null hypothesis that the four groups were drawn from the same population was rejected ($p < .05$; Kruskal-Wallis test).
* Significantly different from dream reporters ($p < .05$).

where I am and aware of waiting for the bell—aware of the hum from the air conditioner"). As one of these *S*s put it, "I'm aware of what I'm thinking about and it's not happening as it would in a dream."

These comments might be summarized by saying that, in the opinion of these *S*s, the self was in an observing role rather than, or in addition to, a participating role in the reported experience. It is particularly interesting to note, therefore, that in the total sample of *S*s, the percentage of reports with the self represented in the narrative was the only content characteristic significantly correlated with the percentage of dream reports given by the *S* ($r = +.56$; $p < .01$). Those *S*s who most frequently labeled their experiences as dreams rather than thoughts gave narratives which more often involved the self as a protagonist of the action. Among the five nonreporters who typically said they were thinking, the self was represented in 40% of the narratives. By comparison, the self was represented in 80% of the narratives obtained from the 12 dream reporters ($p < .01$).

Rem-Period Characteristics of Dream Reporters and Types of Nonreporters

Mean REM-period arousal thresholds and eye-movement activity scores are presented in Table 3 for the dream reporters and for the several subgroups of nonreporters.

As can be seen from the table, the "thinking" nonreporters tended to have lower thresholds than any other group, but they did not differ significantly from the reporters in this respect. When all 46 *S*s were considered, the correlation between the mean arousal threshold for an *S* and the percentage of his awakenings which yielded thinking reports was also in the expected direction ($r = -.20$), but was not significant.

The results were clearer when nonreporters who typically said they were thinking were compared with other *S*s in terms of REM-period eye-movement activity. The thinking nonreporters had significantly less frequent eye movements during their REM periods than did the reporters. For all 46 *S*s, the correlation between mean REM-period eye-movement frequency and the percentage of all awakenings which yielded thinking reports was also significant ($r = -.36$; $p < .05$). In addition, for all 46 *S*s, the REM-period eye-movement score was significantly correlated with the percentage of reports given by an *S* in which the self was represented ($r = +.36$; $p < .05$). Thus the *S*s who had relatively little REM-period eye-movement activity tended to give many reports in which the self was not a protagonist of the action, and which were labeled as thoughts rather than dreams.

A consistent set of relationships therefore emerged that appeared to support the possibility that the experience which presumably occurs during these *S*s' REM periods is somehow different from REM-period experiences in other *S*s, and may actually be more thought-like.

Some of the data on the *S*s who typically said they were thinking seemed paradoxical, however. All but one of these *S*s gave occasional dream reports, but even in these narratives, the self was not often a protagonist of the action, nor was the eye-movement frequency during the REM period which preceded these awakenings higher than usual for the *S*. Although these *S*s tended to have lower eye-movement activity and less self-representation in their narratives, we were unable to predict for any given report whether the *S* would say that it was a dream or a thought. The very dreamlike "thinking" report cited above is an example of a report about which it is impossible at this time to say why the *S* said he was awake and thinking thoughts.

For these "thinking" nonreporters, the possibility should be considered that the pattern of behavior observed in the laboratory was not typical of what it is like at home. One forms the impression that these *S*s may have been particularly unable to relax, fall deeply asleep, and dream in the laboratory. (For example, the only *dream* report given by one of these *S*s was a nightmare in which a black snake was in his bed and got mixed up with the electrode wires.) They might have been inhibiting their dreams in the laboratory as a specific response to what was an anxiety-provoking situation for them.

The group of nonreporters who often recalled dreams from REM-period awakenings were indistinguishable from dream reporters, except for their unusually high arousal thresholds. At this point, it is not clear how this finding helps to account for nonreporting at home, and since the higher threshold was unexpected, there is need for replication and further exploration of this observation.

A third kind of nonreporter is represented by the group of four *S*s who typically gave no-content asleep reports. These *S*s either could not remember that they had been dreaming, or they awakened from REM periods with the impression that they were dreaming but could not recall any content. One *S*, an extreme example of this kind of nonreporter, gave only one line of content in reports from 14 awakenings. Contrary to our starting hypothesis, however, there was no evidence that arousal thresholds were particularly high for these *S*s.

The question of differences between dream reporters and the total nonreporter group was originally one of the major issues guiding the design of this study. Comparisons between these two groups have not been emphasized, however, because the answer to this question may hinge upon an understanding of the nature of the different types of failure to report a dream experience. Since there is the clear possibility that different *S*s are nonreporters for entirely different reasons, it may be misleading to compare dream reporters with the total nonreporter group.[7]

It should be useful in future research to examine the personality characteristics of each of our subgroups of nonreporters, and of the dream reporters. Significant personality differences between dream reporters and nonreporters as a group have been found by several investigators (Lachmann, Lapkin, & Handelman, 1962; Schonbar, 1959; Singer & Schonbar, 1961; Tart, 1962). However, the present study has shown that *S*s who rarely remember their dreams at home are by no means a homogeneous group when studied in the laboratory. Although many unanswered questions remain about the different types of report failure which occurred after experimental awakenings from REM periods, it is evidently useful to distinguish among nonreporters on the basis of their laboratory behavior.

---

[7] In their studies of uninterrupted sleep, Antrobus et al. (1962, 1964) found more eye-movement activity during REM periods among nonreporters than among reporters. In our data, however, no significant differences were observed when the total group of nonreporters was compared with the dream reporters. In view of the many differences between these two studies in procedures and in subjects, it is not possible at this time to pinpoint the reason for the apparent discrepancy in results.

REFERENCES

ANTROBUS, J. S. Patterns of dreaming and dream recall. Unpublished doctoral dissertation, Columbia University, 1962.

ANTROBUS, J., DEMENT, W., & FISHER, C. Patterns of dreaming and dream recall: An EEG study. *Journal of Abnormal and Social Psychology*, 1964, **69**, 341–344.

ASERINSKY, E., & KLEITMAN, N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 1953, **118**, 273.

DEMENT, W., & KLEITMAN, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 1957, **9**, 673–690. (a)

DEMENT, W., & KLEITMAN, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal* *of Experimental Psychology*, 1957, **53**, 339–346. (b)

GOODENOUGH, D. R., LEWIS, H. B., SHAPIRO, A., JARET, L., & SLESER, I. Dream reporting following abrupt and gradual awakenings from different types of sleep. *Journal of Personality and Social Psychology*, 1965, **2**, 170–179. (a)

GOODENOUGH, D. R., LEWIS, H. B., SHAPIRO, A., & SLESER, I. Some correlates of dream reporting following laboratory awakenings. *Journal of Nervous and Mental Disease*, 1965, **140**, 365–373. (b)

GOODENOUGH, D. R., SHAPIRO, A., HOLDEN, M., & STEINSCHRIBER, L. A comparison of "dreamers" and "nondreamers": Eye movements, electroencephalograms, and the recall of dreams. *Journal of Abnormal and Social Psychology*, 1959, **59**, 295–302.

LACHMANN, F. M., LAPKIN, B., & HANDELMAN, N. S. The recall of dreams: Its relation to repression and cognitive control. *Journal of Abnormal and Social Psychology*, 1962, **64**, 160–162.

SCHONBAR, R. A. Some manifest characteristics of recallers and nonrecallers of dreams. *Journal of Consulting Psychology*, 1959, **23**, 414–418.

SINGER, J. L., & SCHONBAR, R. A. Correlates of daydreaming: A dimension of self-awareness. *Journal of Consulting Psychology*, 1961, **25**, 1–6.

SHAPIRO, A., GOODENOUGH, D. R., BIEDERMAN, I., & SLESER, I. Dream recall and the physiology of sleep. *Journal of Applied Physiology*, 1964, **19**, 778–783.

SHAPIRO, A., GOODENOUGH, D. R., & GRYLER, R. B. Dream recall as a function of method of awakening. *Psychosomatic Medicine*, 1963, **25**, 174–180.

SHAPIRO, A., GOODENOUGH, D. R., LEWIS, HELEN B., & SLESER, I. Laboratory studies of dream recall: Gradual arousal as a determinant of the frequency of "thinking" reports. *Psychosomatic Medicine*, 1965, **27**, 342–349.

TART, C. T. Frequency of dream recall and some personality measures. *Journal of Consulting Psychology*, 1962, **26**, 467–470.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: John Wiley, 1962.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received March 4, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 60–64

FAILURE OF AN ALTERNATION–LEARNING TEST TO DISCRIMINATE BETWEEN LOBOTOMIZED AND UNLOBOTOMIZED SCHIZOPHRENICS[1]

RICHARD DE MILLE AND LEONARD LICHT
Brooks Foundation, Santa Barbara *University of California, Los Angeles*

An attempt was made to duplicate Rosvold and Mishkin's finding of clear differentiation between groups, using a modified Yerkes Multiple-Choice Test with 23 lobotomized and 22 unlobotomized schizophrenics for whom scores were available on 10-factored aptitude tests and 2 Wechsler-Bellevue scales. Though marked intergroup differences appeared on the aptitude tests, none appeared on the Yerkes test; score distributions were all but identical, and there was no intergroup difference in perseveration. The Yerkes test was significantly correlated with a Thurstone test. It seemed insensitive to the enduring subtle postlobotomy deficit in schizophrenics demonstrated by factored aptitude tests and in cited studies of alternation learning and multiple-choice learning.

The difficulty of distinguishing between lobotomized and unlobotomized schizophrenics by means of aptitude and intelligence tests is well known (Klebanoff, Singer, & Wilensky, 1954; Lewis, Landis, & King, 1956). De Mille (1962) obtained consistent differentiation using factor-analytic aptitude tests, but most of the differences were not large. Recent experimenters (*E*s) have observed a more striking disparity in performance on alternation and multiple-choice learning tasks.

Pribram, Ahumada, Hartog, and Roos (1964) administered a lever-alternation-learning task to lobotomized and unlobotomized veterans very similar to the subjects (*S*s) in the present study, using mechanical instructions and a small monentary reinforcement. Their lobotomized *S*s made many more simple and complex perseveration errors than their unlobotomized *S*s. Though the sample was very small the effect was so great as to be quite persuasive.

Poppen, Pribram, and Robinson (1965) applied a computer-programed, mechanically-presented multiple-choice test to 22-lobotomized and 18-unlobotomized veterans, using a candy reinforcement. The groups were matched for age, length of hospitalization, education, medication, and degree of psychopathology as rated by nurses; they closely resembled the *S*s in the present study. Marked intergroup differences appeared in overall performance and in several kinds of errors of search and repetition. The high-scoring lobotomized *S*s (those who completed all 20 problems) took many more trials to completion than the high-scoring unlobotomized *S*s.

The most exceptional difference was reported by Rosvold and Mishkin (1950), who applied a modification of the Yerkes Multiple Choice Test (Yerkes, 1921) to lobotomized and unlobotomized schizophrenics and to normals and observed very little overlap between groups. The lobotomized *S*s consistently failed the task; the unlobotomized succeeded most of the time; and the normals always succeeded.

Rosvold and Mishkin's (1950) all male sample consisted of 8 Canadian veterans who had received prefrontal lobotomies 1 to 7 months earlier (*M* 3.4 months),[2] 6-unlobotomized paranoid schizophrenics "of comparable ages, duration of illness, and IQ," and 7 normal 14-year-old boys "of average intelligence and socio-economic background."

No lobotomized *S* could solve all five problems; two solved two problems; the others solved none. Of the unlobotomized schizophrenics, two solved two problems; the others solved all five problems. Each of the seven normal boys solved all of the problems with ease.

---

[1] Supported by a faculty research grant, University of California.

[2] H. E. Rosvold, personal communication.

Rosvold and Mishkin interpreted the observed difference as a specific and enduring result of lobotomy and discussed its implications for treatment. If repeatable, such clear discrimination between lobotomized and unlobotomized would have important implications for contemporary experimental work. The present study attempted to repeat Rosvold and Mishkin's method as closely as possible, with a larger sample of schizophrenics.

METHOD

Subjects

The sample consisted of 23-lobotomized and 22-unlobotomized chronic-schizophrenic white veterans.[3] (An incidental sample of 9 normal *Ss* were tested also; each achieved a perfect score on the Yerkes test, as did Rosvold and Mishkin's normal *Ss*.)

The lobotomized *Ss* had received bilateral prefrontal lobotomies 8 to 14 years earlier (*M* 11.5 years, *SD* 1.9 years). The groups were well matched in age, sex (86% male), type of schizophrenia originally diagnosed, schooling, duration of illness, and whether on a tranquilizer (93%) when given the Yerkes test.[4]

Procedure

Rosvold and Mishkin's description of their procedure was quite brief; additional information was obtained (see Footnote 2) before the following routine was established.

The stimuli were five 3 × 5 blank white file cards, continually replaced so as to reduce the effect of irrelevant cues such as spots or creases.

After the customary social exchange, to establish rapport, *E* and *S* sat face-to-face across a table, alone in a quiet room. The *E*, who did not know which *Ss* were lobotomized, said, "This is a test in which you point out the correct card each time I lay the cards down." The *E* laid out the five cards from his left to right, their 3-inch sides nearest him, and said, "These cards are all alike. One of them is the correct one. I want you to point to the one you think is correct, and if it is the right one, I will say correct. If it is not, I will say wrong. So point to the one that you think is correct."

One trial consisted of laying the cards down, the response of *S*, and retrieving the cards. Success on any of the six problems consisted of 10 correct responses in a row; failure consisted of 50 trials without 10 correct responses in a row. After failure on 2 consecutive problems, testing was terminated. At every trial, the *S*'s response was recorded on a sheet of paper seen only by the *E*.

The transition from one problem to the next was without comment, so that part of the *S*'s task was to recognize that the *E* had changed the card or sequence he was calling correct. Problems I and II were simple position-learning tasks; the other four involved alternation learning. Numbering the cards from *E*'s left, 1-2-3-4-5, the correct responses were as follows: I—The card selected by *S* on the first trial. II—Card 5; 1 if 5 had been chosen in Problem I. III—1 alternating with 3. 1 to be chosen first; then 3 correct until chosen, and so on. IV—5 alternating with 3. V—2 alternating with 4. VI—1-3-5-3.

Success in Problem II, for example, was the series 5-5-5-5-5-5-5-5-5-5; in Problem VI, 1-3-5-3-1-3-5-3-1-3. Problem VI was an addition to the test given by Rosvold and Mishkin, to increase the range of difficulty.

Administration time was about 30 minutes per *S*. Few anomalies occurred in test administration, which was carried out between December 1963, and March 1964.

A summation score and a difficulty score were tabulated for each *S*. The summation score was the number of problems solved; the difficulty score was the index number of the last problem solved; these two were very similarly distributed. The difficulty score was adopted, since its dispersion was greater.

For purposes of correlation, the difficulty score was dichotomized at the point of transition from the simple-position problems to the more difficult alternation problems. In correlations, therefore, this score represents the ability to do the alternation problems.

A perseveration score was tabulated for each *S*. Perseveration was defined as the repetition of a response that had been called wrong on the previous trial. No attempt was made to analyze complex repetitions, infrequently noted in the data, or delayed simple repetitions.

The comparison measures were scores, recorded 3 years earlier, on 10-factored aptitude tests and the Verbal and Performance IQ scales of the Wechsler-Bellevue, Form I. The factored aptitude tests included nine Guilford tests and one Thurstone test (listed in Table 1).

---

[3] Sincere thanks are due the director and staff of Brentwood Veterans Administration Hospital, Los Angeles, in particular Harry M. Grayson, Chief, Psychology Service, for ready cooperation in furnishing subjects.

[4] The subjects were drawn from groups of 55 lobotomized and 56 unlobotomized patients who had been used as subjects in an earlier study (de Mille, 1962). All former subjects who were still available, testable, and acceptable were tested. The attrition in the sample was due to discharge, transfer, or leave from hospital, death, psychological inaccessibility, incomplete data in the original study, elimination of subjects to increase racial homogeneity, and convenience. No systematic differences in reasons for attrition appeared between the groups, except that a small number of the more able lobotomized subjects, who had been tested outside the hospital, were eliminated; their exclusion helped to increase intergroup differences on the comparison tests and presumably favored the appearance of an effect on the Yerkes test similar to that found by Rosvold and Mishkin.

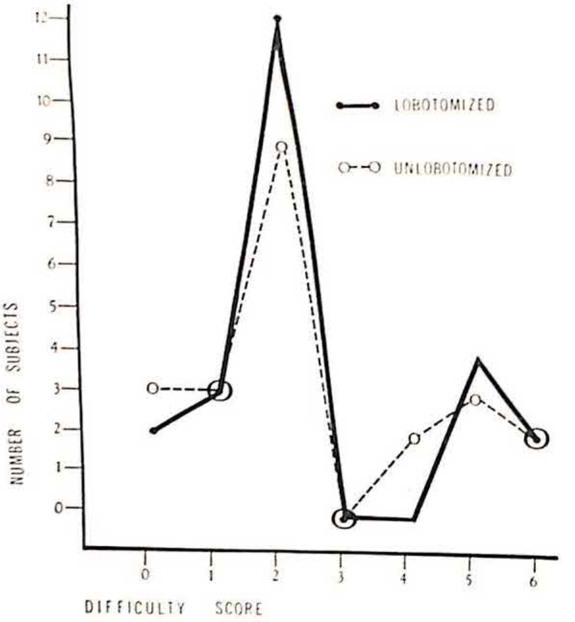


FIG. 1. Distributions of the Yerkes difficulty score for the 23 lobotomized and 22 unlobotomized subjects.

RESULTS

Discrimination of Lobotomized from Unlobotomized

Distributions of the Yerkes difficulty scores for the two groups are shown in Figure 1.[5] The lobotomized mean was 2.5 (*SD* 1.8); the unlobotomized mean was 2.5 (*SD* 1.9).

To test a hypothesis of no difference between the two distributions, chi-square was computed, with the data grouped in three score intervals (0–1, 2–3, 4–5–6). Chi-square equalled 0.56. With two degrees of freedom, the probability that two random samples of data would be more discrepant than the two obtained distributions is at least .70. The null hypothesis was accepted.

The perseveration score, likewise, was similarly distributed for the two groups. The mean number of perseverations per problem attempted was 6.2 for the lobotomized *S*s, 7.6 for the unlobotomized, a nonsignificant difference.

The intergroup differences obtained on the comparison measures are shown in Table 1. As in the earlier study (de Mille, 1962), mean scores were higher for the unlobotomized group on every test. On all but two tests (Verbal Comprehension and Letter Series) the intergroup differences had increased in absolute value with the reduction of the sample. The differences reached the .05 level of significance on 3 of the 10-factored aptitude tests and the .01 level on 3 others. (Significance levels are given for comparison; the outcome was, of course, dependent on the earlier study, where it had already been evaluated statistically.)

TABLE 1

INTERGROUP DIFFERENCES ON APTITUDE AND INTELLIGENCE TESTS

Test	Lobotomized		Unlobotomized		Ratios[a]	
	M	*SD*	*M*	*SD*	*F*	*t*
Numerical Operations	36.83	7.23	50.55	17.53	5.88*	3.38**
Verbal Comprehension	27.65	9.27	29.64	9.11	1.04	.71
Ship Destination	12.39	3.12	18.00	7.57	5.90*	3.20**
Letter Series[c]	7.52	5.98	10.27	5.22	1.31	1.60
Unusual Details	8.70	6.82	11.95	6.66	1.05	1.58
Social Situations	9.70	3.21	11.36	3.34	1.08	1.66
Seeing Problems	6.35	6.99	11.59	8.89	1.62	2.15*
Ideational Fluency	29.74	11.53	40.00	17.27	2.24*	2.30*
Brick Uses (shifts)	1.70	2.14	5.09	5.23	6.00*	2.80**
Alternative Uses	3.61	3.63	6.95	6.28	2.99*	2.15*
Verbal IQ	91.52	11.90	96.77	12.14	1.04	1.43
Performance IQ	91.39	18.26	97.72	16.71	1.19	1.18

Note.—A description of each test is given in the earlier study (de Mille, 1962).
[a] *F* ratio has two-tailed test of significance; *t* ratio has one-tailed test.
[c] An adaptation of Thurstone's PMA Reasoning subtest.
* $p \leq .05$.
** $p \leq .01$.

---

[5] The investigators are grateful to P. R. Merrifield for suggestions about analysis of the data.

Intertest Relationships

As shown in Figure 1, the distribution of the Yerkes difficulty scores for either group is strikingly bimodal, with modes above and below the point of transition from position learning (Problems I and II) to alternation learning (Problems III to VI). Though 7 of the 13 high-scoring *Ss* failed Problem III, the first alternation problem, none failed Problem IV, a fact which strengthens the impression of a true dichotomy reflecting a discontinuity between two kinds of ability. It may be recalled that an equivalent bimodal distribution appeared in Rosvold and Mishkin's unlobotomized group; presumably it did not appear in their lobotomized group only because of the excessive difficulty of the test for that group.

Among the aptitude tests, those which by inspection of four-fold tables appeared related to the Yerkes test were also those whose score distributions were highly skewed or truncated in the sample and presumably in the population (*cf.* de Mille, 1962).

For these reasons, phi (with Yates's correction) was computed between the Yerkes test and each other test and nontest variable for all 45 *Ss*. As expected, all coefficients involving tests or schooling were positive in sign, while those involving age and duration of illness were negative.

By a test of the null hypothesis for each relationship, only Letter Series had a nonzero correlation with the Yerkes test (phi $= .34$, $p < .05$, maximal phi $= .72$). Point-biserial coefficients, where applicable, were not significant.

DISCUSSION

It is clear that in this sample the modified Yerkes test was insensitive to long-range intellectual changes peculiar to lobotomized *Ss*, while factored aptitude tests significantly differentiated lobotomized from unlobotomized. What differences in design between the present study and Rosvold and Mishkin's study can explain the complete negation of their finding?

Rosvold and Mishkin reported matching their two groups of schizophrenics on age, duration of illness, and IQ, but not on schooling or type of schizophrenia. The present investigators discount schooling as a source of bias, since the groups were matched on IQ and the Yerkes test appears little related to academic achievement.

It is not clear whether any bias could have resulted from the use of paranoid schizophrenics in Rosvold and Mishkin's unlobotomized group, since the subdiagnostic composition of the lobotomized group was not reported. Johannsen, Friedman, Leitschuh, and Ammons (1963) found a significant difference between paranoid and nonparanoid schizophrenic veterans on a button-pressing, alternation-learning task, the paranoids scoring higher. The present investigators attribute little importance to such a bias, if it did exist, believing that it could hardly have given rise to the great effect observed by Rosvold and Mishkin.

Duration of illness in the present study was around 18 years; in Rosvold and Mishkin's sample it was around 4 years (see Footnote 2). An interstudy age difference of similar magnitude may be inferred. These differences may help to account for the better performance of Rosvold and Mishkin's unlobotomized group when compared with the present unlobotomized group, but they do not seem particularly relevant to the central finding.

A crucial difference between the studies may be that the mean postoperative interval in Rosvold and Mishkin's sample was 3.4 months, while in the present sample it was 11.5 years. Other investigators have observed marked but transient effects of psychosurgery (Klebanoff et al., 1954; Lewis et al., 1956); the effect observed by Rosvold and Mishkin may have been confined to a short period following the operation. To the present investigators, the interstudy difference in length of postoperative interval offers the most credible explanation of the inconsistency in results.

Evaluation of the Modified Yerkes Test

The failure of the Yerkes test to discriminate between lobotomized and unlobotomized *Ss* must be compared with the successful discrimination between the same groups by factored aptitude tests, and between similar groups by mechanically-presented alternation-learning and multiple-choice tests. The conclusion of the investigators is that the discrepancies in no way disconfirm the aptitude test findings or the recent experimental findings, but do cast doubt on Rosvold and Mishkin's interpretation of their results, and on the modified Yerkes test as a promising instrument for the investigation of frontal-lobe function. It is ironic that the unmodified test (Yerkes, 1921) resembled the more promising laboratory procedures cited above.

The significant correlation between the Yerkes test and the Letter Series test may reflect the fact that each involves a horizontal array of stimuli in which an implicit unfamiliar order is to be discovered. The correlation suggests that the Yerkes test did to some degree measure intellectual performance in the whole sample of 45 schizophrenics and may reflect individual differences among schizophrenics along some intellectual continuum. The subtle additional deficit occasioned by lobotomy in schizophrenics, however, goes undetected by this instrument.

REFERENCES

DE MILLE, R. Intellect after lobotomy in schizophrenia: A factor-analytic study. *Psychological Monographs*, 1962, **76**(16, Whole No. 535).

JOHANNSEN, W. J., FRIEDMAN, S. H., LEITSCHUH, T. H., & AMMONS, H. A study of certain schizophrenic dimensions and their relationship to double alternation learning. *Journal of Consulting Psychology*, 1963, **27**, 375–382.

KLEBANOFF, S. G., SINGER, J. L., & WILENSKY, H. Psychological consequences of brain lesions and ablations. *Psychological Bulletin*, 1954, **51**, 1–41.

LEWIS, N. D. C., LANDIS, C., & KING, H. E. (Eds.) *Studies in topectomy*. New York: Grune & Stratton, 1956.

POPPEN, R. L., PRIBRAM, K. H., & ROBINSON, R. S. Effects of frontal lobotomy in man on the performance of a multiple choice task. *Experimental Neurology*, 1965, **11**, 217–229.

PRIBRAM, K. H., AHUMADA, A., HARTOG, J., & ROOS, L. A progress report on the neurological processes disturbed by frontal lesions in primates. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior*. New York: McGraw-Hill, 1964. Pp. 28–55.

ROSVOLD, H. E., & MISHKIN, M. Evaluation of the effects of prefrontal lobotomy on intelligence. *Canadian Journal of Psychology*, 1950, **4**, 122–126.

YERKES, R. W. A new method of studying ideational behavior of mentally defective and deranged as compared with normal individuals. *Journal of Comparative Psychology*, 1921, **1**, 369–394.

(Received March 16, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 1, 65–70

SPECIFIC AND NONSPECIFIC EFFECTS OF BRAIN LESIONS IN MAN[1]

PAUL SATZ

Clinical Neuropsychology Laboratory, J. Hillis Miller Health Center,
University of Florida

An attempt was made to demonstrate the coexistence of specific and nonspecific effects consequent to the same brain lesions in human Ss. Results indicated that these effects vary, in part, as a function of the tests employed. For example, scores on the Verbal and Performance Scales of the WAIS were shown to be related to specific cerebral areas. When an additional, although more complex, "nonverbal" instrument was employed the results were found to be unrelated to area of lateralization. It was concluded that before attempts to localize "where" effects are involved, one should first proceed to a careful analysis of "what" abilities are measured by the particular tests employed.

Traditional theories of the effects of brain lesions in animals and man are generally dichotomous: that the effects are either specific or general. Meyer (1961) has criticized both theoretical positions for their failure to take adequate account of the conflicting data. The demonstration of concurrent specific and nonspecific effects might therefore advance explanatory alternatives to both hypotheses. It might also provide greater acceptance of the possibility that symmetry and asymmetry of hemispheric function need not represent mutually exclusive concepts in brain function.

Teuber (1959) has suggested that specific effects (e.g., scotomas) can coexist with general effects (e.g., impairment on complex nonverbal perceptual tasks). Somewhat similar findings have been observed in Lashley's work with the rat. Using bilateral ablations, Lashley (1960) demonstrated that habits of the simple conditioned reflex type were dependent only upon the specific sensory areas involved. For more complex tasks, however, such as the conditioned reaction, delayed reaction or multiple stick problems, deterioration occurred after substantial lesions to any part of the cortex. The recent work of Warren, Grant, Hara, and Leavy (1963), using unilateral lesions with rhesus monkeys, lends additional support to the view that specific and nonspecific effects can coexist. An adequate test of this problem would therefore seem to require one lesion to determine the existence of both effects in the same subject. The present study is addressed to this twofold problem.

The first part is concerned with the specific effects of brain lesions, particularly the verbal-nonverbal division of labor as measured by the WAIS (Wechsler, 1958). Factor analytic studies of the earlier Wechsler-Bellevue Scale (Cohen, 1952) found significant loadings on two main factors, a verbal organization and a nonverbal organization factor. Subsequent investigations have indicated that verbal abilities are associated primarily with left cerebral cortex functions, and that nonverbal perceptual abilities are associated with functions mediated by the right cerebral cortex (Milner, 1962; Reitan, 1962). Replication of these findings has not been consistent (Heilbrun, 1956; Meyer & Jones, 1957), nor has methodology always been adequate. In addition, this research has tended to impart considerable surplus meaning to such terms as "verbal" and "nonverbal" when different tests are employed. Finally, there have been few additional attempts to investigate concurrent nonspecific effects.

The second part is focused on possible nonspecific effects of brain lesions and employs an additional more complex measure of "non-

---

[1] This research was carried out while on a National Institute of Mental Health Postdoctoral Fellowship in the Division of Behavioral Sciences, J. Hillis Miller Health Center, University of Florida. The author is grateful to Frederick A. King and Lamar Roberts, Department of Neurosurgery, for their critical comments during the investigation. Particular acknowledgement is extended to my Research Assistant, Eileen Fennell, for her helpful editorial suggestions in preparing the manuscript.

TABLE 1

NUMBER OF PATIENTS IN DIAGNOSTIC CATEGORIES FOR EACH CRITERION GROUP

Diagnostic category	LH	Locus RH	IH
Organic brain cases ($N = 44$)			
Abscess		1	
Cerebral vascular accident	4	1	
Tumors			
Astrocytoma (I,II)	2	3	
Meningioma	1	3	
Closed head injury	1	2	3
Epilepsy	1	3	7
Vascular malformation	1	4	
Cerebral arteriosclerosis	1		4
Cerebral ischema	1	1	
	Frequency		
Medical ward cases ($N = 10$)			
Alcoholism	2		
Cardiac	2		
Headaches	2		
Rheumatoid arthritis	1		
Trigeminal neuralgia	1		
Spinal cord tumor	1		
Seizures, unknown etiology	1		
Psychiatric ward cases ($N = 24$)			
Schizophrenic reaction	12		
Reactive depression	4		
Anxiety reaction	5		
Sociopathic personality	2		
Adolescent adjustment reaction	1		

Note.—LH = Left Hemisphere, RH = Right Hemisphere, IH = Indeterminate Hemisphere.

verbal" perceptual performance. Previous investigation has found performance on this complex multivariate test (Block Rotation Test) to be generally independent of locus, laterality or type of brain lesion (Satz, 1963, 1964). This test, however, did demonstrate rather powerful discrimination between organic and nonorganic criterion groups regardless of age, IQ, and education.

METHOD

Subjects

Subjects were chosen from in-patient wards of the Teaching Hospital, University of Florida, School of Medicine. All subjects were given complete physical examinations. On the basis of these exams, subjects were divided into four groups: (*a*) left-hemisphere lesions ($N = 12$), (*b*) right-hemisphere lesions ($N = 18$), (*c*) indeterminate-nonlateralized lesions ($N = 14$), and (*d*) medical and psychiatric controls ($N = 34$). The three neurological groups were carefully evaluated by the Neurology and Neurosurgery Services of the hospital. Special neurological tests included skull films, brain scans, electroencephalography, arteriography, and pneumoencephalography. Approximately 40% of the neurological subjects ($N = 17$) underwent surgery for treatment of neoplastic and atrophic brain lesions. The diagnostic categories for the neurological and control groups are presented in Table 1. The vascular, neoplastic, and atrophic brain lesions were all confined to either the left- or right-hemisphere group. Approximately 70% of the indeterminate-hemisphere group, however, was composed of the nonfocal traumatic head-injury cases and idiopathic convulsive disorders.

Table 2 presents the mean data on age and Full Scale IQ (FIQ) for all groups. Analysis of variance of FIQ scores revealed sizable differences among groups on intellectual level ($F = 5.15$, $p < .01$), but no significant difference among groups for age.

TABLE 2

MEAN SCORES ON WAIS FULL SCALE IQ AND AGE FOR EACH GROUP

	LH	RH	IH	C
IQ	99.67	103.78	88.14	98.79
Age	45.50	41.00	40.29	42.10

Tests

Specific effects. WAIS Verbal (VIQ) and Performance (PIQ) Scale scores were employed as measures of verbal and nonverbal performance.

Nonspecific effects. The Block Rotation Test, a 44-item visual-motor test designed for measurement of the rotation effect, was employed to test generalized cortical function. In the administration, the subject is presented with a series of simple block designs which are constructed by the experimenter. The designs are composed of two, three, or four blocks and involve solid red and white colors only. The subject is required to reconstruct each design (given additional blocks) as it would look if rotated 90 degrees to the left or to the right. The test is divided into two parts: Part A consists of 15 stimulus designs presented in either the vertical or horizontal position; Part B consists of 7 stimulus designs presented at a 45 degree angle to the vertical or horizontal axis. A total of six rotation error variables are employed in the scoring; these errors are: errors in Part A, errors in Part B, total errors, duplication errors, angulation errors, and time errors. Age and Performance IQs are also included in the scoring.

The rotation error variables were selected on the basis of their significant correlation with brain dysfunction. Age was included because of its significant correlation with test errors. Performance IQ was used because of its high inverse correlation with errors and as a control for the nonverbal nature of this task (Satz, 1963, 1964).

This procedure introduces a novel test situation for the subject which minimizes the use of previous experience and/or learning. The Block Rotation Test (BRT) differs further from other tests of brain function by requiring more than a mere reproduction or recall of the stimulus materials. The rotated solution to each design is not embedded in the subject's stimulus field, but rather, is thought to depend heavily upon the utilization of more complex brain-perceptual processes.

Discriminant function analysis (Fisher, 1936) was employed as the multivariate statistical procedure to predict the presence or absence of cortical dysfunction using eight continuous variables. This function is expressed by the following linear equation:

$$Z = \lambda_1 X_1 + \lambda_2 X_2 + \ . \ . \ . + \lambda_n X_n$$

in which Z is the composite predictor score based on the individual scores for each of the variables or tests (X_1, X_2, . . . X_n), and on the respective weights assigned to each of these scores (λ_1, λ_2 . . . λ_n). The weights used were based on the values arrived at in a previous discriminant function analysis of this test (Satz, 1963).

Hypotheses

The use of differential WAIS Scale scores generated a variety of hypotheses, only three or four of which have traditionally been investigated.

Specific effects (within groups). The following four hypotheses were tested:

I. Verbal performance (VIQ) would be lower than nonverbal performance (PIQ) in cases of left-hemisphere lesions.

II. Verbal performance (VIQ) would be higher than nonverbal performance (PIQ) in cases of right-hemisphere lesions.

III. Verbal performance (VIQ) would be no different than nonverbal performance (PIQ) in cases of indeterminate-hemisphere lesions.

IV. Verbal performance (VIQ) would be no different than nonverbal performance (PIQ) in the control cases.

Specific effects (between groups). The following two hypotheses were tested:

V. Verbal performance (VIQ) in the left-hemisphere cases would be lower than verbal performance (VIQ) for the right-hemisphere, indeterminate-hemisphere, and control cases.

VI. Nonverbal performance (PIQ) in the right-hemisphere cases would be lower than nonverbal performance (PIQ) in the left-hemisphere, indeterminate-hemisphere, and control cases.

Nonspecific effects. Chi-square analyses were run on the distribution of discriminant function Z scores to test the following hypotheses:

VII. The Z scores for the combined neurological groups and for the controls would *not* be independent of the two criteria (organic versus control).

VIII. The range of Z scores between the neurological groups would be independent of the criteria of classification (i.e., that the Z scores would not be related to the site of the lesion).

RESULTS

Tests for specific effects (within groups). Results of this analysis (Table 3) show that all four hypotheses were clearly supported, both in direction and in magnitude. The differential scale score of 9.92 IQ points ($t = 2.67$, $p < .05$) for the left-hemisphere group shows rather striking impairment in verbal cognitive functioning (VIQ). In marked contrast is the nonverbal perceptual impairment on PIQ (7.44 IQ points) for the right-hemisphere group ($t = 2.47$, $p < .05$). Neither the indeterminate-hemisphere nor the control group, however, shows a significant difference between VIQ and PIQ ($t = 1.52$, $p > .20$).

TABLE 3

MEAN DIFFERENCES BETWEEN VERBAL AND PERFORMANCE IQ FOR EACH GROUP: TEST FOR SPECIFIC EFFECTS (WITHIN GROUPS)

Group	VIQ	PIQ	$\overline{D}$	$S_{\overline{D}}$	t
LH	94.92	105.17	9.92	3.72	2.67*
RH	107.22	99.78	7.44	3.01	2.47*
IH	90.21	87.00	3.29	2.16	1.52
C	100.03	97.09	2.94	2.21	1.33

* $p < .05$.

Tests for specific effects (between groups). It was felt that an additional, if not more critical test of the verbal-nonverbal division of labor would be to treat each criterion group as an independent treatment group being measured on only one WAIS Scale at a time. In view of support for the first four hypotheses, one would expect maximal impairment on Verbal IQ functions for the left-hemisphere group (Hypothesis V), and maximal impairment on Performance IQ functions for the right-hemisphere group (Hypothesis VI). A basic problem encountered in such an analysis relates to the marked difference in Full Scale IQ (FIQ) between the four treatment groups. Covariance adjustments for FIQ were therefore made on VIQ and PIQ separately.

The results of these two covariance analyses (Hypotheses V & VI) are presented in Tables 4 and 5, respectively.

Inspection of Table 4 indicates an overall difference between groups on the adjusted Verbal IQ scores ($F = 6.10$, $p < .01$). Individual comparisons of the adjusted mean scores show that the lowest VIQ was obtained by the left-hemisphere group (VIQ = 93.46).

TABLE 4

ANALYSIS OF COVARIANCE OF VERBAL IQ SCORES FOR EACH GROUP WITH FULL IQ CONTROLLED: TEST FOR SPECIFIC EFFECTS (BETWEEN GROUPS)

Sources	$C_{YY'}$	*df*	$ms_{Y'}$	*F*	$F_{.99}$
A	523.60	3	174.53	6.10*	4.08
w	2,087.49	73	28.60		

Differences between adjusted group means on VIQ[a]

Group	RH	IH	C
LH	−8.32**	−6.48*	−5.97*
RH		1.84	2.35
IH			.51

[a] Adjusted VIQ group means are as follows: LH = 93.46, RH = 101.78, IH = 99.94, C = 99.43.
* $p < .01$.
** $p < .001$.

This adjusted mean was significantly lower than that of each of the remaining three groups ($p < .01$), which would lend support for Hypothesis V.

The results of tests of Hypothesis VI are presented in Table 5. Covariance analysis indicates an overall difference between treatment groups on the adjusted Performance IQ scores ($F = 4.45$, $p < .01$). When the differences between each of the PIQ means are compared, however, the results indicate only partial support for this hypothesis. The adjusted mean PIQ was indeed lower when compared with the other three groups, but the only expected difference occurred between the left- and right-hemisphere groups ($t = 4.35$, $p < .001$).

TABLE 5

ANALYSIS OF COVARIANCE OF PERFORMANCE IQ SCORES FOR EACH GROUP WITH FIQ CONTROLLED: TEST FOR SPECIFIC EFFECTS (BETWEEN GROUPS)

Sources	$C_{YY'}$	*df*	$ms_{Y'}$	*F*	$F_{.99}$
A	672.48	3	224.16	4.55*	4.08
w	3,595.21	73	49.25		

Differences between adjusted group means on PIQ[a]

Group	RH	IH	C
LH	9.21**	7.51*	7.26*
RH		−1.70	−1.95
IH			−.25

[a] Adjusted PIQ group means are as follows: LH = 103.77, RH = 94.56, IH = 96.26, C = 96.51.
* $p < .01$.
** $p < .001$.

Tests for nonspecific effects. The discriminant function z scores were divided into four categories, with two falling above the critical cutoff point ($Z = 450.05$) and two falling below it. A chi-square test of significance was run between the criteria of organic (LH, RH, IH cases) versus control (medical and psychiatric cases) and the range of Z scores. The computed chi-square value of 32.285 was highly significant (df 3, $p < .001$), thus supporting the hypothesis of a relationship between z scores and criteria of brain disorder versus no brain disorder (Hypothesis VII). The cutoff value ($Z = 450.05$) correctly classified 82.05% of the subjects. Errors in classification were as follows: false positives, 17.65%, and false negatives, 18.18%.

A second chi-square test of significance was run between left-hemisphere, right-hemisphere, and indeterminate-hemisphere groups and the distribution of Z scores. The computed chi-square value of 5.31 was not significant (df 6, $p > .50$), thus supporting the hypothesis of independence between the Z scores and the site of the lesion in the organic cases (Hypothesis VIII).

DISCUSSION

The present findings suggest that the verbal-nonverbal division of labor, *as measured by the WAIS*, does provide one meaningful framework for discussion of hemispheric asymmetry of function in man. Such findings are, of course, not new. The discrepant findings of Heilbrun (1956) and Meyer and Jones (1957) do present serious counterevidence to the assumed dominance of nonverbal-perceptual functions for the right cerebral cortex. One plausible explanation for these discrepancies is that different "nonverbal" measures were employed in the above studies. Considerable caution would therefore seem warranted, particularly in extrapolation, when employing different nonverbal measures.

A discrepancy in nonverbal function was obtained in the present study by employing an additional more complex perceptual task (BRT). Under this condition, the behavioral effects were found to be unrelated to the site of hemispheric lesions, thereby giving support to the hypothesis which predicted nonspecific effects with this task. Similar nonspecific effects were obtained in a previous study by Teuber and Weinstein (1956) employing complex visual and tactual tasks. For example, visual performance on the Embedded Figures Test (Teuber, 1959) was found to be unrelated to either locus or laterality of brain lesions. The *complexity* of the dependent variables employed was cited by Lashley (1963) to account for possible general effects of lesions in rats. He reasoned: "that the maze habits are more complex, are dependent upon the activity of larger amounts of tissue, and hence are not functionally identified with any particular cerebral area [p. 141]." According to Teuber and Weinstein (1956): "If adequate performance on a test depends on a number of different functions each of which is crucial, then any lesion sufficient to upset one of them may lead to significant general impairment [p. 377]."

If behavioral effects, after cerebral lesions, are shown to vary so closely as a function of the nature of the tests employed, then much more attention to these factors would seem necessary before further attempts are made to delineate specialized centers in the animal and human brain. The attempt to ascribe test deficits to specific cerebral lesions without an analysis of the abilities involved in test performance is hazardous (Orbach, 1959). Teuber and Weinstein (1955) have pointed out that all too often the tendency is to stop when the "where" is demonstrated, without proceeding to analyze "what" we mean by localization. According to Meyer (1961): "If the nature of a deficit is not precisely defined by careful investigation, the only valid conclusion is that a lesion produces some complex dysfunction resulting in impairment on a test [p. 56]." With this in mind, the following analysis of performance on the Block Rotation Test (BRT) is presented in an attempt to account for the nonspecific effects with this test.

A number of possible functions relating to performance on the BRT have previously been discussed (Satz, 1963). The subject is not asked to merely copy stimulus lines or drawings, nor is he asked to remember these designs. Furthermore, he is not required to find hidden figures in geometric patterns. Correct performance seems to require a displacement from the sensory stimulus cues, and the activation of more complex brain-perceptual processes, that is, the ability to organize perceptually the figure-ground relationships of the various designs in the absence of a visual representation of the final arrangement.

Three postulated levels of performance on the BRT were separately analyzed in an earlier study (Satz, 1963). Level I was defined as the perception of the stimulus design as presented by the experimenter. Level II was defined as perception of the rotated stimulus image; and Level III was defined as motoric translation of the rotated perceptual image. These three levels apparently interact in that an impairment at Level I, that is, an inability to correctly perceive the stimulus design, leads to a breakdown in performance at the two higher levels. In like manner, an error at Level II, that is, the inability to correctly visualize the rotated stimulus design, necessarily leads to an error in performance at Level III. An impairment at Levels II or III does not, however, necessitate an error at Level I; for the subject may be able to correctly perceive the initial stimulus design, but may not be able to perform correctly at the two more difficult levels (Levels II and III). Evidence for this test behavior was found in the previous study in which several of the organic subjects failed by merely reproducing the experimenter's stimulus designs. A reproduction of the stimulus design, however, indicates by definition, correct perception at Level I. This type of error performance parallels rather closely the concept of "stimulus boundness" cited by other investigators (Goldstein, 1959; Hemmendinger, 1953). It also suggests that different levels of perception might be involved in performance on a task such as the BRT; first, a lower level recognition-discrimination system, and second, a higher level system involving more complex integrative perceptual processes. Bortner and Birch (1960, 1962) have recently advanced support for a similar hypothesis of levels in perception.

The preceding analysis of the BRT indi-

TABLE 1

ROTATED FACTOR MATRIX OF ANTISOCIAL BEHAVIOR ITEMS

Antisocial Behavior Items[a]	Factor I	Factor II	h²
1. Driving a car without a driver's license	.339	.047	.117
2. Theft (less than $2.00)	.339	−.519	.384
3. Damaging or destroying private or public property	.549	−.226	.352
4. Truancy	.562	−.159	.341
5. Premarital and extramarital sex relations	.546	−.070	.303
6. Cheating on school examinations	.315	−.016	.099
7. Theft (between $2.00 and $50.00)	.524	−.447	.474
8. Getting drunk	.416	−.002	.173
9. Smoking against parents' wishes	−.146	−.530	.302
10. Defying parents' authority to their face	.202	−.639	.449
11. Disobeying parents behind their back	.081	−.515	.272
12. Exceeding speed limits	.226	−.261	.119
13. Use of profane or blasphemous speech	.279	−.417	.252
14. Theft (taking things from someone's locker)	.019	−.267	.072
Σa^2	1.890	1.820	3.709
Total Variance	13.5	13.0	26.5

[a] The following scoring system was used: never = 0, once or twice = 1, several times = 2, and often = 3.

of sexual behavior, and Nye and Short (1957; Short & Nye, 1958) in their study of juvenile delinquency. Nye and Short found that certain antisocial activities form a Guttman type scale (1957). This scale, however, cannot be readily used with adult subjects.

In the present study the factor analytic approach was used in order to construct an antisocial behavior scale. The subjects were administered a 14-item antisocial behavior inventory consisting of appropriate items from the Nye and Short studies plus additional items. Some of these items refer to antisocial activities specific to childhood and adolescence; others to adult life. Items referring to parental disobedience were included because of the theoretical relevance of this variable for antisocial behavior. The subjects were requested to indicate how frequently (never, once or twice, several times, and often) they have engaged in each of the various antisocial activities. The subjects were assured about the anonymity of their responses, and were encouraged to give honest replies.

TABLE 2

MEAN AND STANDARD DEVIATION SCORES OF FATHER-ABSENT AND FATHER-PRESENT GROUPS ON THE *ASB* AND *PD* SCALES

Group	*ASB*-scale *M*	*ASB*-scale *SD*	*ASB*-scale *t*	*PD*-scale *M*	*PD*-scale *SD*	*PD*-scale *t*
Father-Absent	5.69	2.77	1.98*	3.40	1.66	<1.00
Father-Present	4.76	2.47		3.46	2.34	

* $p < .05$.

The responses of 53 randomly selected students were factor analyzed, in order to determine whether there are distinct clusters of antisocial behavior. The factors were extracted by the principal-axis method (Harman, 1960) and subjected to a varimax rotation.[2]

In order to ascertain whether assurances concerning anonymity eliminated the "social desirability" response set, all subjects were given Edwards' (1957) and Christie's social desirability (*SD*) scales. The subjects' questionnaires were coded, which made it possible to retest them about 3 months later and to obtain reliability estimates.

RESULTS AND DISCUSSION

Table 1 presents the results of the factor analysis of the antisocial behavior inventory. Only factors with eigenvalues larger than 1.0 were rotated. Two factors were obtained: Factor 1 suggests an antisocial behavior factor, and Factor 2 a parental disobedience factor. Items 3, 4, 5, 7, and 8 were selected to form an antisocial behavior (*ASB*) scale, and Items 9, 10, and 11 were selected to form a parental disobedience (*PD*) scale. The test-retest (*r*) reliability of the ASB scale is .84, and that of the *PD* scale .83.

The correlations (*r*) between subjects' *ASB* scale and their Edwards *SD* scale and Christie *SD* scale scores were −.14 and −.10, respectively. The correlations (*r*) between subjects' *PD* scale and their Edwards *SD* scale and Christie *SD* scale scores were −.08 and −.11, respectively. None of the above correlations are significant, suggesting that when subjects receive adequate assurance about the anonymity of their replies the "social desirability" response set is not a significant source of variance.

Table 2 indicates that the father-absent group obtained significantly higher *ASB* scale scores than the father-present group, but that there was no significant difference between the two groups in relation to the *PD* scale. Thus, the results of the present study are consistent with the hypothesis that father-absence is associated with antisocial behavior.

---

[2] Table A, presenting the correlation matrix and the *ASB* and *PD* scales, have been deposited with the American Documentation Institute. Order Document No. 8713 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. Remit in advance $1.25 for microfilm or $1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service.

In a study of 6–10-year-old children, Bach (1946) found that the father fantasies of father-absent boys were more similar to those of girls than were the fantasies of father-present boys. Sears, Pintler, and Sears (1946), in a doll-play study of preschool children, found less aggressive doll play among the boys of a father-absent group than among the boys of a father-present group. These findings suggest a more feminine orientation in the boys of the father-absent groups than in the boys of the father-present groups. On the other hand, there is also evidence that as father-absent boys grow older they tend to react to their feminine identification with exaggerated masculinity. Thus, Lynn and Sawrey (1959) who studied 8–9½-year-old children of Norwegian sailors—who are absent from home for intermittent but extended periods of time—noted insecure father identification with considerable compensatory masculine behavior. The relationship between father-absence and antisocial behavior obtained in the present study is consistent with the findings obtained by Lynn and Sawrey.

The findings of the factor analysis have two obvious implications for the psychology of moral behavior. First, the emergence of an antisocial behavior factor is in clear opposition to the conclusion reached by Hartshorne and May (1928) that the violation of moral standards or, conversely, the conformity to moral norms, is highly situational and shows no evidence of consistency from one moral situation to another. It should be pointed out, however, that the kind of violation which Hartshorne and May were primarily concerned with, that is, dishonesty on examinations, had only a moderate loading on the antisocial behavior factor. Thus, it may be that *some* categories of antisocial activity are indeed highly situational. It should also be pointed out that the subjects in the Hartshorne and May studies were younger than the participants in the present study, and it is possible that generalized antisocial response tendencies only emerge with maturation.

The second point concerns the emergence of two orthogonal factors, one consisting of antisocial behavior and the other essentially of parental disobedience. This finding suggests that not *all* types of antisocial behavior can be conceptualized in terms of defiance of social authority which in turn is related to an earlier defiance of parental authority. On the other hand, as can be seen in Table 1, items Number 2 and 7, which refer to theft, do have moderate loadings on both factors. Moreover, item Number 13, referring to the use of profane and blasphemous speech, has a significant loading only on the parental-disobedience factor. Thus, there is evidence that *some* forms of antisocial behavior are in fact related to defiance of parental authority. If antisocial activities tend to form a limited number of identifiable clusters, as is suggested by the results of the factor analysis, it is indeed reasonable to assume that such clusters or factors will be related to different antecedent variables. Therefore, it is not surprising that defiance of parental authority is related to some but not all forms of antisocial behavior. However, further investigation of both these issues, that is, the patterning of antisocial behavior and their causal antecedents, is clearly indicated.

REFERENCES

Bach, G. R. Father-fantasies and father typing in father separated children. *Child Development,* 1946, **17**, 63–80.

Bacon, H. K., Child, I. L., & Barry, H. A. A cross-cultural study of correlates of crime. *Journal of Abnormal and Social Psychology,* 1963, **66**, 291–300.

Cohen, A. K. *Delinquent boys: The culture of the gang.* Glencoe, Ill.: Free Press, 1955.

Edwards, A. L. *The social desirability variable in personality assessment research.* New York: Dryden Press, 1957.

Harman, H. H. *Modern factor analysis.* Chicago: Univer. Chicago Press, 1960.

Hartshorne, H., & May, M. A. *Studies in deceit.* New York: Macmillan, 1928.

Kinsey, A. C. *Sexual behavior in the human-male.* Philadelphia: W. B. Saunders, 1948.

Lynn, D., & Sawrey, W. L. The effects of father absence on Norwegian boys and girls. *Journal of Abnormal and Social Psychology,* 1959, **59**, 258–262.

NYE, F. I., & SHORT, J. F., JR. Scaling delinquent behavior. *American Sociological Review,* 1957, **22,** 326–331.

PARSONS, T. Certain sources and patterns of aggression in the social structure of the western world. *Psychiatry,* 1947, **10,** 172.

SEARS, R. R., PINTLER, M., & SEARS, P. S. Effect of father separation on preschool children's doll play aggression. *Child Development,* 1946, **17,** 219–243.

SHORT, J. F., JR., & NYE, F. I. Criterion of deviant behavior. *Social Problems,* 1958, **5,** 207–214.

(Received March 19, 1965)

ERRATUM

The footnote identification in Table 6 (page 62) of the article "The Problem of Validity in Field Studies of Psychological Disorder," by Bruce P. Dohrenwend and Barbara Snell Dohrenwend (*J. abnorm. Psychol.*, 1965, **70**, 52–69), should have appeared as follows:

TABLE 6

MINIMUM AND MAXIMUM RATES OF PSYCHOLOGICAL DISORDER REPORTED ACCORDING TO SOCIOECONOMIC STATUS

	Percentage		d	Author(s)
	Minimum	Maximum		
	Minimum in highest stratum			
Maximum in lowest stratum	19.5	27.0	7.5	Bremer[c], 1951
	1.3[a]	3.2[a]	1.9	Akimoto et al.[b], 1942
	0.7 1.1	3.7 6.6	3.0 5.5	Cohen et al.[d], 1939
	30.0	37.8	7.8	Taylor & Chave[i], 1964
	7.3	16.6	9.3	Hyde & Kingsley, 1944
	17.4[a]	29.5[a]	12.0	Bellin & Hardt[c], 1958
	1.6 6.0	15.1 25.4	13.5 19.4	Gnat et al.[j], 1964
	12.5	47.3	34.8	Srole et al., 1962
	—	—	—	Cole et al.[e], 1957
	—	—	—	D. Leighton et al.[f], 1963
	Minimum in a middle stratum			
	2.3[a]	2.9[a]	0.6	Tsuwaga et al.[b], 1942
	0.8	1.9	1.1	Lin, 1953
	1.2[a]	2.5[a]	1.3	Brugger[g], 1931
	10.3	29.7	19.4	Primrose, 1962
	Minimum in lowest stratum			
Maximum in a middle stratum	3.4[a]	5.3[a]	1.9	Brugger[g], 1937
	45.0	54.1[a]	9.1	Llewellyn-Thomas[h], 1960
	7.4[a]	25.7[a]	18.3	Brugger[g], 1933
	Minimum in a middle stratum			
Maximum in highest stratum	6.2	13.6	7.4	Pasamanick et al., 1959

[a] Calculated by B. S. Dohrenwend.

[b] Distribution of population in socioeconomic strata was reported only by number of families; since Japanese census reports do not include information on family size by SES, rates were calculated on the assumption of equal family size in all four strata. Rates for socioeconomic strata are below total rate for Tsuwaga due to reduction by 12 of number of cases reported according to SES.

[c] Subjects divided into only two strata.

[d] Data for wh'tes only reported for two wards separately.

[e] Cole et al. do not report rates but state: "Four-fifths of the families in the lower social strata contained at least one mentally ill member, while less than one-half of the upper-stratum families were thus affected [p. 395]."

[f] Results reported in ridits rather than percentages.

[g] Occupations grouped by B. S. Dohrenwend into three strata: high (self-employed merchants, manufacturers, and farmers, and middle level civil servants, middle (merchants, manufacturers, and farmers employed by others, and low level civil servants), and low (workers and servants).

[h] Occupations grouped by B. S. Dohrenwend into three strata: high—independent business and salaried workers, middle—fisherman and farmers, low—laborers. The maximum rate in these strata is below the overall rate because of the exclusion of two categories of persons with high rates, grouped separately by Llewellyn-Thomas: i.e., housewives, with a rate of 76%, and miscellaneous, with a rate of 65%.

[i] Based on survey data without supplementary physicians' reports included in total rate.

[j] Rates for two cities reported separately.

"*In the field of international social psychology,* THE PATTERN OF HUMAN CONCERNS *is clearly a work of first magnitude.*"
—PROFESSOR GORDON ALLPORT, *Harvard University*

The Pattern of Human Concerns

by HADLEY CANTRIL

Using a brilliantly original technique of psychological exploration, this simply stated book proposes a new way of looking at mankind's social structure. Employing what they term the Self-Anchoring Striving Scale, Dr. Cantril and his associates reveal how people everywhere feel about their lives and nations, their pasts, presents, and futures.

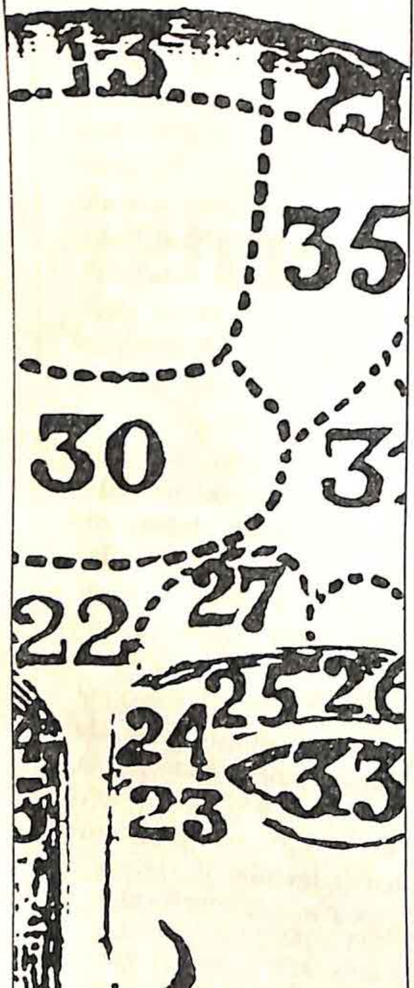
Their study embraces nearly a third of all the people on earth—citizens of both developed and developing countries, of democracies, of dictatorships—a coverage of fourteen nations, five continents, and many islands, including—importantly—Cuba and the Dominican Republic. Dr. Cantril's fresh method of approach shows how it is possible to avoid many errors in the exercise of national policy, and to embark on policies which offer a likelihood of success.

Dr. Cantril has been an adviser to four of the last five United States administrations. His report deals with the real world. Its implications embrace politics, government, history, social science, economics, cultural anthropology, and a dozen other fields.
Charts, diagrams, 450 pages, $10.00

This volume is one of a group of major works published in celebration of the two hundredth anniversary of the founding of Rutgers, The State University.

RUTGERS UNIVERSITY PRESS, New Brunswick, New Jersey

McGraw-Hill books in PSYCHOLOGY



PRINCIPLES OF PSYCHOPATHOLOGY: An Experimental Approach

By Brendan A. Maher, University of Wisconsin. McGraw-Hill Series in Psychology. Off press.

Oriented towards research documentation and the scientific method, this book applies principles of scientific investigation to abnormal behavior. The treatment of the subject matter brings abnormal psychology into the realm of psychology. It derives its approach from the principles of general psychology. The text covers the areas of behavior designated by the terms *neuroses*, *psychosis*, and other *special deviations*.

THE DISORGANIZED PERSONALITY

By George W. Kisker, The Graduate School, University of Cincinnati. Text: 656 pages, $8.50. Taped case interviews: $60.00 (complete set).

This introductory text in abnormal psychology contains both the fundamentals of formal, experimental abnormal psychology, and a fresh, clinical orientation based on the author's first-hand experience as a clinical practitioner in the field. Of particular importance is the inclusion of selected case histories, accompanied by, and keyed to, recorded interviews.

ABNORMAL PSYCHOLOGY: Understanding Behavior Disorders

By Jack R. Strange, Southern Methodist University. 500 pages, $7.95.

A concise, complete, introductory text designed for students with no previous knowledge of the field. The approach is "psychobiosocial"—stressing the fact that all behavior results from the interaction of many physical, mental, and interpersonal factors. Includes theories of personality, description and etiology of the behavior disorders, as well as methods of therapy and research.

HANDBOOK OF CLINICAL PSYCHOLOGY

By Benjamin B. Wolman, Adelphi University, in collaboration with seven consulting editors: Gordon F. Derner, Molly Harrower, Robert R. Holt, James G. Miller, O. Hobart Mowrer, Henry A. Murray, and Silvan S. Tomkins. 1596 pages. Handbook Edition $32.50. Text Edition for classroom adoption or individual purchase of college instructors: $24.50.

A rigorously scientific, objective, and systematic work covering all aspects of clinical psychology. Represents clinical psychology as a science, whether it applies to scientific research, teaching, or service.

Send for your examination copies today
McGraw-Hill Book Company
330 West 42nd Street, New York, N.Y. 10036

The 1965 Directory of the American Psychological Association

A STANDARD REFERENCE: a listing of the 23,500 Members of the Association

BIOGRAPHICAL DATA: on Fellows, Members, and Associates

A VALUABLE TOOL: for psychologists, and for

MEMBERS OF RELATED PROFESSIONS
LIBRARIES
PUBLISHERS
PERSONNEL OFFICES

PRICE: $10.00
MEMBER PRICE: $5.00

Order Form

American Psychological Association
Order Dept.,
1200 17th St., N. W.
Washington, D. C. 20036

Date————

Please send——copies of the APA Directory to:

————————
(please print or type)
————————
————————

Check enclosed———— Please bill————

Signature————

A PENETRATING ANALYSIS OF HUMAN BEHAVIOR IN THIS NEW TEXT FROM MACMILLAN

An Introduction to Social Psychology

by Richard Dewey, University of New Hampshire, and W. J. Humber, Consulting Psychologist, Milwaukee

A comprehensive introduction to social psychology, this text is designed to provide the student with a clear conception of the nature of human behavior. It emphasizes a special interactionist frame of reference as the central phenomenon of human behavior; personality is viewed as the result of the interaction of inherited or physiological factors as well as sociological and psychological ones. Parts One and Two expose the student to theory, then demonstrate applications of the concepts to personality, personality types, and to various forms of group action. The Third Part treats theoretical applications to age groups, from infancy through old age. This material is rarely available in a text of this kind. Actual case studies, drawn from the authors' own professional experiences, illustrate the theories and their applications, as do the numerous photographs and line drawings.

1966, 656 pages, $7.95

Write to the Faculty Service Desk for an Examination Copy

THE MACMILLAN COMPANY
60 FIFTH AVENUE, NEW YORK 10011

Oxford University Press

Perspectives in Psychopathology:

Readings in Abnormal Psychology

Edited by **JAMES O. PALMER** *and* **MICHAEL J. GOLDSTEIN**, University of California, Los Angeles

Consisting of twenty-eight important, up-to-date research papers in psychopathology, this work concentrates upon areas and methods of study rather than disease entities. Perspectives of the science of abnormal behavior form the basis for organization of the readings, with the first half of the book centering on three methodological perspectives, the Clinical, the Actuarial, and the Experimental. The papers in the second half of the work shift to an etiological or causal emphasis, illustrating the Biological, Socio-Cultural, and Developmental perspectives. In a thorough introduction to each major section, the editors define the perspective and discuss its advantages and limitations. A pamphlet of questions to accompany each selection will be made available to teachers.

1966 430 pp. 44 illus. paper $4.00

The Experience of Anxiety: *A Casebook*

By **MICHAEL J. GOLDSTEIN** *and* **JAMES O. PALMER**, University of California, Los Angeles

These twenty-four unanalyzed cases are designed to introduce the student to the more important theories of psychopathology, to review the relevant research on abnormal behavior, and to provide a variety of illustrative case materials. "Excellent case histories, well written and organized, with questions well designed to teach the student the process of drawing conclusions from clinical material."—Clifford H. Swensen, Purdue University.

1963 304 pp. paper $2.25

Principles of Group Treatment

By **ERIC BERNE**, M.D., Lecturer in Group Therapy, Langley-Porter Neuropsychiatric Clinic, San Francisco

In this first systematic treatise on the use of transactional analysis in groups, an attempt is made to mold group treatment into a more rigorous and less empirical discipline. The work emerges as the result of Dr. Berne's more than twenty years of experience in practicing psychotherapy. It includes chapters on group dynamics, teaching, research and writing, and the literature of group treatment, always from the viewpoint of the practicing clinician with his special problems.

March 1966 approx. 320 pp. $6.50

Foundations of Psychopathology

By **JOHN C. NEMIAH**, M.D., Harvard Medical School and Massachusetts General Hospital

1961 338 pp. $6.50

Oxford University Press / 417 Fifth Ave. / New York, N.Y. 10016

Volume 71 April 1966 Number 2

Journal of Abnormal Psychology

Editor: Howard F. Hunt, *New York State Psychiatric Institute and Columbia University*

Associate Editor: William N. Thetford, *College of Physicians and Surgeons, Columbia University*

Consulting Editors

Theodore X. Barber
Medfield Research Foundation

Harold Basowitz
New York University

Eugene I. Burdock
New York State Psychiatric Institute and Columbia University

John M. Butler
University of Chicago

Arthur C. Carr
New York State Psychiatric Institute and Columbia University

Loren J. Chapman
Southern Illinois University

John H. Cumming
New York State Department of Mental Hygiene

W. Grant Dahlstrom
University of North Carolina

Jerome D. Frank
Johns Hopkins University

Sol L. Garfield
Columbia University

Ann Magaret Garner
Lincoln, Nebraska

Israel Goldiamond
Institute for Behavioral Research

Harrison G. Gough
University of California, Berkeley

Starke R. Hathaway
University of Minnesota

Ralph Heine
University of Chicago

Ernest R. Hilgard
Stanford University

Lawrence C. Kolb
New York State Psychiatric Institute and Columbia University

Sheldon J. Korchin
University of California

William G. Lawlor
Fordham University

Ardie Lubin
U. S. Naval Hospital, San Diego

Salvatore Maddi
University of Chicago

David Rosenthal
National Institute of Mental Health

Alan K. Rosenwald
Illinois State Psychiatric Institute

Saul Rosenzweig
Washington University, St. Louis

Jacques Rutschmann
New York State Psychiatric Institute

Seymour Sarason
Yale University

Joseph Schachter
New York State Psychiatric Institute and Columbia University

Roy Schafer
Yale University

John Shlien
University of Chicago

Morris I. Stein
New York University

Charles Strother
University of Washington

Harold L. Williams
University of Oklahoma School of Medicine

C. L. Winder
Michigan State University

Joseph Zubin
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR	FRANCES L. BREWER	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Editorial Assistant	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1966 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 75–86

RORSCHACH MOVEMENT AND DREAMS:

A VALIDATION STUDY USING DRUG-INDUCED DREAM DEPRIVATION[1]

BARBARA LERNER[2]

University of Chicago

To test the hypotheses that dream deprivation would produce an increase in quantity of projected movement and a change in its quality in the direction of body-dissolution imagery, 20 Ss were tested with Holtzman inkblots before and after 2 nights of drug-induced dream deprivation. Their performance was compared with that of 30 non-dream-deprived controls. Results were significant in the predicted direction and were interpreted as offering support for Rorschach's views on the fundamental similarity between movement and dreams due to the centrality of kinesthetic experience in both. Additional findings on space were congruent with Fonda's view of space as resistance to internal, rather than external, forces. More specifically, in the present study, space was interpreted as resistance to movement.

Herman Rorschach felt that there was a crucial similarity between movement responses and dreams in that kinesthetic experience in the absence of overt motor activity was central to both. Support for part of this formulation exists. Beginning with the work of Singer, Meltzoff, and Goldman (1952) and Meltzoff, Singer, and Korchin (1953), a great deal of evidence has accumulated to show that Rorschach was correct about the relationship between motor inhibition and the movement response (Bendick & Klopfer, 1964; Goldman & Herman, 1961; Neel, 1960; Singer & Herman, 1954; Singer & Opler, 1956; Singer & Spohn, 1954). Evidence for a similar relationship between motor inhibition and dreams was produced by Mourly Vold (1910–12) and, more recently, by Dement and Wolpert (1958). Thus, support exists for the relationship between motor inhibition and movement responses and motor inhibition and dreams. To date, however, there has been no experimental evidence on the similarity Rorschach assumed to exist between movement responses and dreams or on the importance of kinesthetic experience in dreams.

This study is an attempt to secure such evidence by using a new avenue opened by recent experimental work on the dream process and its physiological concomitants. Using the electroencephalogram and the electro-oculogram, Aserinsky and Kleitman (1953, 1955) and Dement and Kleitman (1957a, 1957b) have shown that it is possible to determine, with a high degree of reliability, when the subject is dreaming. It has been found that subjects are extremely consistent in the amount of time spent dreaming each night. When this uniform nightly quantity of dream time is experimentally reduced by either repeated awakenings or suitable drugs, the dream-deprived subjects compensate for their loss at the first available opportunity. On postdeprivation nights, dream time increases by as much as 50% and remains high until there has been a rough quantitative compensation for the amount of dream time lost. Several studies (e.g., Dement, 1960; Fisher & Dement, 1963; Rechtschaffen & Maron, 1964; Williams, Hammack, Daly, Dement, & Lubin, 1964) have now established this deprivation-compensation effect quite dramatically, and the inference seems clear: dreaming has psychic relevance apart from sleep, and exists as an independent phenomenon, fulfilling a vital psychic need which increases if unmet and is compensated for at the first available opportunity.

---

[1] This study was completed in partial fulfillment of the requirements for the degree of doctor of philosophy at the University of Chicago. The work was aided, first by funds from PHS S01 FR 05369 administered through the University of Illinois College of Medicine and later by NIMH grant 1 FL MH-25,283-01 administered through the University of Chicago. The author is indebted to NIMH and to the following persons for their very considerable help: William Henry, Donald Fiske, Melvin Sabshin, Ernest Haggard, Joseph Wepman, Margery Baittle, and Karol Weinstein.

[2] Now at Mental Health Center, Chicago, Illinois.

On the basis of these findings, it was hypothesized that if movement and dreams are really similar processes, dream-deprived subjects who are tested before having an opportunity to sleep normally will attempt to compensate for their loss by projecting a significantly greater quantity of movement responses onto the inkblots, thus offering some evidence that the hypothesized relationship exists.

To explore the role of kinesthesia in that relationship, specific qualitative changes in movement were also predicted. These predictions were derived from the following tentative understanding of the relationships between kinesthesia, movement, and dreams. Movement responses and dreams are basically kinesthetic experiences and, as such, have their roots in the body-image phenomenon.[3] One of the functions of dreaming is to reintegrate[4] body image and fantasy and, when the opportunity to do this is lessened by dream deprivation, a dissolution of body image begins to take place and is reflected in the quality of movement produced. Accordingly, the second major hypothesis of this study is that dream-deprived subjects will show a significant change in quality of movement responses in the direction of body-dissolution imagery, reflecting a disturbance in body image as a result of dream deprivation and indicating the importance of kinesthetic elements in both movement and dreams.

METHOD

The subjects were 50 graduate and undergraduate students from Chicago area universities and professional schools. All subjects were paid volunteers between the ages of 17 and 28, reportedly in good health and using no drugs other than the research ones.

Dream deprivation was achieved by the administration of 15 milligrams of d-amphetamine sulphate in spansule form plus 100 milligrams of pentobarbital sodium at bed time. This type of drug-induced dream deprivation has been shown to reduce dream time by approximately one-half with the amphetamine as the chief agent and the pentobarbital augmenting its dream-depriving effects while counteracting its sleep-depriving ones (Dement, Clemes, Keating, & Ruster, 1964; Fisher & Dement, 1963; Oswald, Berger, Jaramillo, Keddie, Olley, & Plunkett, 1963; and Rechtschaffen & Maron, 1964).

To eliminate problems arising from inter-subject differences in preexperimental base rate of movement, an own-control design was utilized, with a test-retest difference score for each subject serving as the measure of change in movement. Because test-retest with the original 10 Rorschach inkblots poses innumerable problems, the inkblots chosen for this study are from the Holtzman series. A split-half, rather than an alternate form selection, was made because Holtzman, Thorpe, Swartz, and Herron (1961) report that Form A pulls more movement than Form B. Thus, to assure a movement-equivalent stimulus, an adequate sample of movement responses, and a relatively constant number of R,[5] the test consisted of the first 22 odd-numbered Holtzman A cards and retest consisted of the first 22 even-numbered Holtzman A cards. Each subject was instructed to give two responses per card on both test and retest. All testing was done by the investigator, and inquiries were standardized as much as possible by adhering to a consistent set of prearranged limits.

Subjects were divided into three groups: (*a*) experimental, $n = 20$, 10 male and 10 female; (*b*) placebo control, $n = 20$, 10 male and 10 female; (*c*) drug control, $n = 10$, 5 male and 5 female. Experimental subjects were tested in the morning, then dream-deprived by taking the drugs at bed time that night and the following night, then retested the next morning. The same procedure was followed for placebo controls, except that they were given placebos instead of active drugs. In addition, an attempt was made to institute a check on effects of the actual drugs other than their dream-depriving ones. Accordingly, drug-control subjects were given the same amphetamine dose as experimental subjects, but during waking hours to permit measurement of its effects, if any, on movement when its dream-depriving effect was ruled out. These subjects were tested in the evening (after 5 PM) and told to take their medication on arising for the next two mornings, then retested in the evening of the second drug day. The drug-control group is, however, an imperfect one for two reasons: first, considerations of safety made it necessary to give the amphetamine alone, without the barbiturate for daytime use and, second, it is possible that the amphetamine spansule had not been completely metabolized by the time some of these subjects went to bed. Hence, they may also have experienced dream deprivation but to a considerably lesser extent. All subjects were given their drugs and instructions on when to take them after the first testing session and were asked to omit daytime napping and the use of alcohol and drugs other than the research ones during the entire 3-day period of their participation in the study.

---

[3] "Image" is perhaps a poor word; it suggests simple visual representation or abstract ideation rather than an integrated kinesthetic experience of self as the basis of ego which is the sense in which it is intended here.

[4] Reintegration is necessary because the bulk of an adult's inner life is neither acted out nor experienced kinesthetically except in dreams, except under conditions of severe pathology.

[5] *Relative* rather than *absolute* because (*a*) often, what a subject called one response contained two or more scorable formulations and (*b*) several subjects occasionally gave more than two responses even by their own definition, for example, after giving two R, adding, "Oh! what it really looks like now is a . . . ." These "extra" responses were accepted and tallied as long as they were offered spontaneously during the free association period. "Extras" are to be distinguished from "additionals," the latter being responses which first arose during the inquiry. Additionals were noted but not included in any of the tallies.

HYPOTHESIS I: CHANGES IN QUANTITY OF MOVEMENT DURING DREAM DEPRIVATION

To test Hypothesis I, a sensitive, highly reliable and meaningful method of scoring was desired. Two problems presented themselves in this context. First, interscorer reliabilities for the Rorschach as a whole tend to be only moderately high, ranging from 69 to 86 percent agreement (Voigt & Dana, 1964). Second, Rorschach experts differ on what should and should not be scored for movement. In an effort to deal with both of these problems, two a priori decisions were made. First, it was decided to simplify the scoring task by separating the problem of scoring movement from other scoring problems: only movement was scored. Second, a special movement scoring system was devised in an attempt to take advantage of the ideas of both Beck (1961) and Klopfer and Kelly (1942) and to do so in a way that would improve interscorer reliability and facilitate comparisons between systems and categories of movement without the necessity for tedious rescoring.

Movement was defined as a combination of affect and impulse which, for the moment, is being directed inward and lived out, bodily, in fantasy. It is not so much activity per se as it is empathic kinesthesia; hence, there is passive as well as active movement, and intensity of experience, not quantity of motion, is the salient criterion.[6] What is central is the experience of bodily sensation, the sense of what it would feel like to be in a certain position, to take a certain emotional stance or to engage in a certain activity—anything from lying perfectly still and limp to running, glowering, dancing, etc. Hence, any action, stance, facial expression, or kinesthetic experience, whether in a human, animal, or inanimate representation, and whether active or passive, is movement. However, since kinesthesia seems most likely to be experienced when movement is projected onto a stimulus construed to be like the self and clearly described in kinesthetic terms, the following weighting system was designed.

Weight of 2. The weight of two is assigned to unequivocal human movement, whether active (running), passive (looking), or in between (walking).

To animal movement which Beck would score M (animals engaged in distinctively human movement or with human motivation or in human dress).

To very clear and forceful inanimate or abstract movement.

Weight of 1. The weight of one is assigned to equivocal human movement, defined as responses which describe conventional positions only, without expressive stance or even minimal activity, and where evidence for postural kinesthesia is not explicit, for example, "people standing," "people facing each other," "people sitting," "people lying"—without further embellishment—for example, "people lying, relaxed and limp" would be weighted 2. Occasionally, if the context in which such equivocal responses occur is strongly suggestive of affective stance, a 2 scoring is given, for example, "person sitting alone at the bottom of an abandoned mine shaft."

To clear animal movement which is not distinctively human, for example, "two wolves leaping at each other" or "two dogs looking at each other."

To clear human or animal movement embedded in an art concept, for example, "painting of a man running in fear" or "statue of an angry horse rearing up on his hind legs." Where the art concept appears to be brought in only secondarily, to explain away some detail not relevant to the scoring of movement, the response should be weighted as if the art context had not been introduced.

---

[6] Holtzman's M scoring system (Holtzman et al., 1961) was rejected because it makes the opposite assumption, weighting M responses in terms of "energy level" in the sense of physical energy required to perform the action if it were done overtly.

To inanimate or abstract movement which is clearly present but does not meet the criterion for 2.

Weight of 0. The weight of zero is assigned to equivocal animal movement, for example, animals "facing each other," "sitting," "lying," or "standing"; bird, bat, butterfly, or airplane "flying," "in flight," or "with outstretched wings," and fish, duck, or other water fowl "swimming." These latter are assumed to be conventional descriptive phrases rather than empathic kinesthesias, unless additional evidence strongly indicates otherwise.

To art responses in which posture is noted without enlivening attitude or activity, for example, "idols squatting" or "statue with arms outstretched" is weighted 0, but "idol squatting morosely" or "statue with arms outstretched, beckoning" is weighted 1. (With further embellishment, these responses could achieve a weight of 2.)

To equivocal inanimate movement responses, for example, atom bomb explosion responses which appear to be determined solely by the conventional mushroom-shaped cloud and/or color and shading.

RESULTS

Because distribution of Rorschach variables like movement tends to be skewed (Fiske & Baughman, 1953), and because the assumption of equal scale units cannot be made (Cronbach, 1949), a decision was made to treat the data ordinally throughout, and statistical techniques were chosen accordingly.

All 100 records (test and retest on 50 subjects) were scored for movement by two independent scorers[7] using the above-described scoring and weighting system. The total for each record was called the weighted total movement (WTM) score and the difference between each subject's test and retest WTM, his WTM-D score, was used as the measure of change in movement. Spearman's rank-order correlation was used to measure reliability in scoring WTM-D with a resultant r_s of .96, a highly satisfactory result ($p >$ .001).

To ascertain the comparability of the experimental and control groups before experimental manipulation, mean age and initial WTM scores were computed for each group. As can be seen from Table 1, the two crucial groups, experimental and placebo, are extremely well matched on these factors. Matching was somewhat less successful in the drug-control groups: these subjects tend to have a lower initial WTM score than either of the other two groups.

The basic hypothesis, that dream deprivation results in increased projection of movement, was tested by comparing WTM-D scores in the experimental group with those of each control group on the basis of a Mann-Whitney *U* test. The hypothesis was strongly supported: experimental subjects have significantly higher WTM-D scores than either placebo or drug-control subjects. Results in the first comparison are significant at the .001 level and in the second at the .05 level (see Table 4). These results seem to indicate that dream-deprived subjects do attempt to compensate for their loss by projecting more movement responses onto inkblots. Thus, they suggest some degree of functional equivalence between the processes involved in dreams and movement responses.

It seemed possible, however, that these results might be deceptive in their specificity;

TABLE 1

MEAN AGE AND INITIAL WTM SCORES FOR EXPERIMENTAL AND CONTROL GROUPS

Group	$\bar{x}$ Age	$\bar{x}$ WTM	*SD*	Range
Experimental $n = 20$, 10 M and 10 F	21.8	35.45	15.61	5–64
Placebo control $n = 20$, 10 M and 10 F	20.8	37.40	16.07	17–65
Drug control $n = 10$, 5 M and 5 F	21.6	22.00	12.84	8–49

---

[7] The two scorers were the experimenter and a colleague, Karol Weinstein. All of Weinstein's scoring was completely blind while the experimenter's scoring was blind in only about half the cases. To avoid bias in the results, the experimenter's scoring was used only to ascertain interscorer reliability. All comparisons between experimental and control groups were made on the basis of Weinstein's scoring only, except for one instance where a third scorer was brought in to make additional comparisons.

perhaps the increase in WTM as a result of dream deprivation merely reflected an overall increase in productivity, a rise in R generally rather than in movement particularly. To test this possibility, a third and very experienced blind scorer[8] was given the 10 experimental cases that showed the greatest increase in WTM and asked to do full scoring on all 20 records (test and retest on each subject), using the present scoring system for movement and Beck's system for other determinants. Two of these 10 cases were found to show no rise in R on retest. The remaining eight showed a rise in both R and WTM. Rises in both areas were then compared by computing WTM-D as a percentage of R-D (response total difference score). Because WTM is a weighted score whereas R is not, R-D was also compared with the absolute total movement difference score (TM-D). From the results shown in Table 2, it seems clear that the rise in R is largely a function of the rise in movement and not the reverse. Thus, the specificity of the relationship between Rorschach movement responses and dreams was further substantiated.

However, the meaningfulness of this substantiation in the context of the other results reported here is dependent on the comparability of WTM scoring in each instance. This was assessed by comparing the third scorer's WTM-D scoring with that of both previous scorers on the basis of Spearman correlations. Scorer three and the investigator achieved an r_s of .78, and scorer three and the other scorer achieved an r_s of .81. Results are significant at beyond the .01 level in both cases, and this is particularly encouraging in light of the fact that the 10 cases involved constitute a small homogeneous subsample with a sharply restricted range of variation.

TABLE 2

RELATION OF MOVEMENT TO RESPONSE TOTAL IN EXPERIMENTAL SUBJECTS WITH THE TEN HIGHEST WTM-D SCORES

Case number	Percent increase in R attributable to an increase in WTM	Percent increase in R attributable to an increase in TM
M16	89	54
F13	100	100
F7	100	100
M17	77	50
M18	100	76
F15	80	49
M8[a]	—	—
M3	40	21
M4	81	62
F1[a]	—	—

[a] These two *S*s showed no rise in R.

These results indicate that a real increase in movement production took place in the experimental group. However, because the drug-control group has several previously cited imperfections, one might still wonder if the increase in movement might not somehow be a function of drug intake rather than of dream deprivation per se. In other words, if dream deprivation was achieved through a method which did not involve drug intake, would movement still show a significant increase? Relevant evidence on this point is available from two previous studies of sleep deprivation, which is, of course, a form of dream deprivation. In the first, Loveland and Singer (1959) found an increase in M% on the Rorschach significant at the .01 level after 100 hours of sleep deprivation. Because these authors had not hypothesized this result in advance and had no a priori explanation for it, and because it was one of only 3 out of 100 different test scores which showed changes in central tendency significant at the .05 level or better, they dismissed it as probably due to chance. This "chance" finding, however, showed up again in 1963 in Palmer's investigation of stress. One of his stress situations was 120 hours of sleep deprivation, and the subjects in this group showed "a marked intratensive shift" (1963, p. 213) on the Holtzman as measured by the formula $EB = 100 \left[\frac{M - \text{Sum } C}{R}\right] + 100$. The specificity of this result to sleep deprivation (dream deprivation) is highlighted by comparison with his other two "stress" groups for whom the stressor was food deprivation. Subjects in the first food-deprivation group showed no change, and subjects in the second showed a definite change in the opposite direction, a shift toward extratension.

In light of these comparable findings on dream deprivation without drugs, it seems unlikely that the increase in movement in dream-deprived subjects in the present experiment was a result of drug intake rather than of dream deprivation.

---

[8] Margery Baittle.

TABLE 3

A COMPARISON OF CATEGORIES OF MOVEMENT BASED ON ABILITY TO DISCRIMINATE BETWEEN DREAM-DEPRIVED AND NON-DREAM-DEPRIVED GROUPS

Group	Category	Sample size	Mann-Whitney U	p
Experimental versus Placebo control	M-D	20 × 20	117	.02
	FM-D	20 × 20	143.5	.10
	m-D	20 × 20	182.5	*ns*
Experimental versus Drug control	M-D	20 × 10	73	*ns*
	FM-D	20 × 10	81	*ns*
	m-D	20 × 10	98.5	*ns*

Findings thus far indicate that a real change in quantity of movement took place in the experimental group and that this change was a result of dream deprivation. However, because WTM is a global, weighted score, it does not reveal the precise nature of the change in terms of absolute numbers or particular categories of movement responses. To clarify the contribution of each of the three major categories of movement in common use to this overall result, separate test-retest difference scores were also computed for M, FM, and m. For this purpose, weighted scores were abandoned and the absolute number of responses in each category was relied upon. Results presented in Table 3 show a clear progression: in comparisons between the experimental and the placebo-control groups, M alone discriminates most finely (although not at as high a level of significance as WTM), FM shows a significant trend, and m fails to discriminate at all. In comparisons between the experimental and the drug-control group, although none of the three categories taken in isolation discriminates at a significant level, the same order of progression is evident, as can be noted from the size of the Mann-Whitney U for each category.

These results suggest that Beck's scoring system, which applies successively more stringent criteria for scoring movement in human, animal, and inanimate movement, in that order, might make a somewhat finer discrimination between groups than either the present WTM system or Klopfer's system, since the present system has more lenient standards for FM and m than Beck's, while Klopfer's is more lenient yet. Such comparisons between systems are possible on the basis of the present weighting system. One can approximate Beck's scoring by simply counting the absolute number of movement responses weighted 2, and one can similarly approximate Klopfer's scoring by taking the absolute number of all movement responses, those weighted 1 as well as 2.

In this manner, separate test-retest difference scores were computed for each case on the basis of both systems. Mann-Whitney U tests were then run on the results of each and compared with those previously described for WTM-D, to learn which of the three systems discriminated most sharply between the groups. To clarify further the relationships involved, separate counts of the absolute number of only those movement responses weighted 1 were also made for each case, and significance tests were made on this basis also. From the results presented in Table 4, it can be seen that all three systems give highly significant results in experimental versus placebo-control comparisons while the weight of one component alone does not. In experimental versus drug-control comparisons, where even finer discrimination is required because some drug-control subjects probably experienced some slight amount of dream deprivation, only the Beck and Lerner systems give significant results. Moreover, in both types of comparison, the same order of progression is apparent: Mann-Whitney U is lowest (hence best) for Beck's system, higher for Lerner's, and highest for Klopfer's. Taken together, the nonsignificant results achieved with responses weighted 1 alone and the nonsignificant results achieved with m alone seem to account for the relative superiority of Beck's system to either the author's or Klopfer's in the present research. Since Beck's system is the most faithful to Rorschach's original conception,[9] the results might

---

[9] Considerable confusion exists about Rorschach's conception of movement. For example, it is often asserted, quite erroneously, that only human movement is movement in Rorschach's (1951) sense and that he excluded the possibility of movement in nonhuman representations. In fact, he specifically included it: "There are some subjects who can perceive movement not only in human figures and animals with certain human characteristics but in all kinds of animals, plants, geometric figures, and even in single lines [p. 26]." What he makes clear on the preceding page is that the likelihood of movement decreases when nonhuman content is perceived,

TABLE 4

A Comparison of Movement Scoring Systems Based on Ability to Discriminate between Dream-Deprived and Non-Dream-Deprived Groups

Group	System	Sample size	Mann-Whitney *U*	*p*
Experimental versus Placebo control	Weights of 2 only (Beck)	20 × 20	71	.001
	Weighted Total Movement[a] (Lerner)	20 × 20	85	.001
	Absolute Total Movement (Klopfer)	20 × 20	104	.005
	Weights of 1 only	20 × 20	190	*ns*
Experimental versus Drug control	Weights of 2 only (Beck)	20 × 10	55	.025
	Weighted Total Movement[a] (Lerner)	20 × 10	60.5	.05
	Absolute Total Movement (Klopfer)	20 × 10	71.5	*ns*
	Weights of 1 only	20 × 10	93.5	*ns*

[a] This is the only weighted score in the table. All other system scores are based on absolute number of responses, e.g., absolute number of movement responses weighted 2, absolute number of all movement responses whether weighted 1 or 2, and, last, absolute number of movement responses weighted 1.

be interpreted as highlighting Rorschach's precision of focus as well as his creative originality in inventing and defining the concept of movement.

In this context, it might be appropriate to note that in the eight previously cited studies on the part of Rorschach's formulation concerning the relation between motor inhibition and movement, several report more significant results with M than with FM and m while only two report the reverse: Neel (1960) obtained more significant results with m, and Goldman and Herman (1961) obtained more significant results with FM.

Debates about the relative merits of different categories of movement are nonetheless likely to continue for some time since a number of ambiguities still exist and the issue is by no means resolved. In spite of this, the overall trend seems clear: some form(s) of movement was found to increase following motor inhibition or to be related to inhibiting capacity in eight studies, and some form(s) of movement was found to increase following dream deprivation in three studies, including the present one. Together, these quantitative shifts lend strong support to the totality of Rorschach's concept of movement.

In the present study, qualitative findings provide further support for the quantitative ones already cited. Movement responses given by experimental subjects on retest were often strikingly dreamlike in quality and, in at least one case, appeared to verge on the hallucinatory, as in the following example (subject is a nonpsychotic engineering student who showed the highest rise in WTM in the experimental group. The response is to Card 12A):

> This one's going to be rather difficult to explain. There's a number of birds around the base of a tree. *There's a—not in this picture, but—somewhere else,* there's a picture of Christ knocking at the door. His robe is open and the sacred heart shows. This tree also has a heart and in that heart is an insect, white, crawling towards the top. The roots of the tree are like fingers with thumbs sticking out, forwards.

This subject's test lasted approximately 90 minutes and produced nothing even remotely similar to the above response. His retest lasted for more than 4 uninterrupted hours, during which he vehemently rejected the suggestion that he might like to take a break or get some food. He gave every impression that the opportunity to project movement onto the blots was fulfilling an urgently experienced need, one which superseded even his hunger as he worked long past his usual lunch time. Other experimental subjects showed a similar intensity and absorption following dream deprivation. One interesting manifestation of this was what came to be called "the last chance upsurge" phenomenon, a striking increase in productivity on the last few retest inkblots, as if these subjects were loath to see the experiment end. Other qualitative changes are discussed in relation to Hypothesis II.

Hypothesis II: Changes in Quality of Movement during Dream Deprivation

To provide a basis for an initial, partial test of this hypothesis, a scoring system was devised to take account of breakdowns in body image as expressed in movement responses. This was done by assigning all move-

---

hence, more evidence is required to substantiate its presence and stricter scoring standards must be applied.

ment responses to one of the six categories described below. Three of these categories (M, FM, and m) are variants on Klopfer's (Klopfer & Kelly, 1942) distinction between human, animal, and inanimate movement. The other three were devised to take account of various types of blurring, fusion, and confusion between M, FM, and m which seem to reflect body boundary weakness at the conceptual level. In addition, a secondary category, body dissolution (BD), was used for responses which seemed to reflect body-image dissolution at a concretely tactile level, for example, when there is not only a dissolution in the experiencer but when the experience itself is a sense of body dissolution. A movement response in any of the six categories may receive the additional designation of BD if it meets the criteria described below.

M Human movement: assigned to all responses where movement is projected onto what is basically a human form whether it is a whole human form (H) or part of a human form (Hd or Hdx). Religious, mythological, and cartoon figures with a basically human shape are also categorized as M, for example, angel, devil, satyr, centaur, ghost, Martian, tin woodsman, wolf man, and most monsters. Also, humans in costumes representing other species, for example, "Man in a bear costume dancing."

Mb Human movement, borderline:[10] assigned to all responses where movement is projected onto what is basically a human form but in an inanimate yet enlivened representation, for example, "Mount Rushmore in a conversational mood—those carved faces appear to be talking," "a wooden puppet, dancing, he's moving very jerkily," "skeleton of a man clinging to this object for dear life."

FM Animal movement: assigned to all responses where movement is projected onto what is basically an animal form whether whole (A) or partial (Ad or Adx) and whether real, for example, bear, horse, dog, or imaginary, for example, Pegasus, Dumbo, unicorn. Movement in an animal form is categorized as FM even when the animals are clearly humanized, for example, "old lady mice in aprons, holding hands," "elephants riding motorcycles," "birds sneering contemptuously at each other."

FMb Animal movement, borderline:[11] assigned to all responses where movement is projected onto what is basically an animal form but in an inanimate yet enlivened representation, for example, "rocking horse struggling to break free."

m Inanimate movement: assigned to all responses where movement is projected onto a form which is neither human nor animal, for example, "volcano erupting, shooting up with tremendous force," "wind blowing very hard, the movement there, whipping up, this way," "a venus fly trap snapping shut on this speck," "abstract forces pulling away from the center which is buckling."

Mfm Confabulated movement: assigned to confabulated movement responses in which there is a confused and/or shifting combination of various types of movement, and which are not scorable in terms of any of the above categories. For example:

"And in the background, I see a ghostly mask, one of those masks that just covers the eyes. Seems to be looking at them, peering at them with some evil design."

> Inquiry: Center piece over bridge of nose, slits for the eyes and general part that would cover around the eyes. Something evil associated with it. I guess it's the position of the eyes, slanted downward, squinting. Reminds me of cat's eyes. Funny thing is, we have two cats and I adore them, but cats are kind of evil, lurking pets of witches. "Cat" in that sense. These are dark and I think of a black cat's eyes, peering, scrutinizing these men. Men are unaware. They're looking at each other and the cats are looking at the men. 'Course, it's not really a cat but a mask, but I have the impression there's some kind of power, force behind the mask with evil design. Even our cats aren't so nice, at night, when they get outside, hunting, carnivorous, no longer pets. But they're nice cats.

This extensive example is quoted in full in the hope of providing a clear illustration of the type of genuinely confabulated response which belongs in this category, as distinguished from the merely fanciful or literary combinations and/or alternatives which many intelligent and well-educated subjects produce and which can usually be categorized under one of the preceding five headings.

---

[10] Borderline not with respect to the presence or absence of movement, but to its categorization as human (M) or inanimate (m).

[11] See Footnote 10 for the sense in which borderline is meant, in this case, between animal (FM) and inanimate (m) movement.

BD Body dissolution: affixed at the end of the scoring formula for movement responses in which the body is in the process of dissolution *without* a causative external agent. Included here are responses in which the body loses its form definition or boundaries and "melts" or in which body parts become separated or in which internal organs become separate and outside the body. This scoring is intended for kinesthetic imagery which seems to reflect a disturbance in basic body image arising out of experiences of the body which may be thought of as existing on a continuum between somatic depersonalization and autoscopia.

Some of the situations in which these deviant experiences of the body seem likely to arise follow: in dreams and during sexual orgasm; with prolonged sleep deprivation (dream deprivation); in the induction phase of a hypnotic trance; in some drug and sensory-restriction experiments; in acute anxiety attacks; in schizophrenic disorganization; and in certain organic states. Although anxiety is a frequent accompaniment of these somatic experiences, they are distinguished from other anxiety-invoking bodily experiences not by the presence of pain but by the absence of ordinary sensations of physical boundaries, integration, stability, wholeness, and familiarity. They are likely to be described as strange, unusual, or indescribable rather than painful, and they range from a pleasurable sense of merging with one's surroundings during orgasm to a terrifying feeling of complete dissociation from one's physical self in incipient schizophrenia.

Examples of Rorschach movement responses which would be scored for body dissolution are:

> That almost looks like two eyes, a face, looking out at the world and this is sort of the rest of him which has come loose and he's wondering where it is and these seem to be bugs, attacking it. Yet, he doesn't look too sad, mostly curious.
>
> Inquiry: Eyes first caught my attention, looking, not with pain or distress or quite as strong as dismay either. Maybe, in a way, anger. Face (pointing), rest of person, bugs, head and legs, shape of feelers in an up position (demonstrates: "clawed" hands). Now, not so much attacking but eating, and the eyes look like they're watching but they won't go down there. Mouth down here but came apart.

There are two movement responses in the preceding example: the looking and wondering face that has lost its body is a clear body-dissolution image and would be scored DM H/BD. The attack of the bugs is not a BD response and would be scored simply DFM A.

An earlier and milder, but still recognizable, stage of body dissolution is reflected in the following response:

> This one looks like a face, in a way, again, with a skeletal quality, and the lines leading up to what looks like eyes. Seems to have sort of a light but the edges are pressed out. Seems like it's trying not to see, and feet are somehow leaded down.
>
> Inquiry: Feet (pointing), no clearly defined shape of muscle, almost like a hoof. No. Like the expression "Your Gods have feet of clay." Eyes somehow looking into the light, trying to avoid looking into this fuzz.

Scored: DMFm · Confab/BD.

RESULTS

Each record was given one point for every response categorized as Mb, FMb, or MFm and an additional two points for every response scored BD. The total number of points derived in this manner was called the weighted movement dissolution (WMD) score. Each subject's test and retest WMD score was then compared to produce a single weighted movement dissolution difference score (WMD-D) for each case. The reliability of these WMD-D scores was assessed by comparing the blind scorers' results with an independent scoring by the investigator on the basis of Spearman's rank-order correlation. The resulting r_s, after corrections for ties, was .75, indicating an acceptable level of reliability.

To ascertain the comparability of the three groups prior to dream deprivation, mean initial WMD scores were computed for each group. Results showed WMD scores to be initially closely comparable and very low in all three groups, averaging less than one per record. The hypothesis that dream deprivation produces a disturbance in body image as reflected in movement responses was then tested by comparing the WMD-D scores in the dream-deprived group with those in the two non-dream-deprived groups on the basis of Mann-Whitney U tests. Results were significant in the predicted direction (see Table 5), indicating that dream deprivation does tend to produce a dissolution in body image which is reflected in an increase in movement-dissolution responses projected onto inkblots.

TABLE 5

DIFFERENCES BETWEEN EXPERIMENTAL AND CONTROL GROUPS ON WMD-D SCORES AND ON TWO TYPES OF SPACE SCORES

Group	Sample size	Variable	Mann-Whitney *U*	*p*
Experimental versus Placebo control	20 × 20	WMD-D[a]	96	.005
		PS-D[b]	98.5	.005
		SS-D[c]	179.5	*ns*
Experimental versus Drug control	20 × 10	WMD-D	54	.025
		PS-D	54.5	.025
		SS-D	61.5	.05

[a] Weighted movement dissolution difference score.
[b] Primary space difference score.
[c] Secondary space difference score.

A POST HOC HYPOTHESIS: USE OF SPACE DURING DREAM DEPRIVATION

The above findings were made on the basis of a priori hypotheses and scoring systems derived from a theoretical understanding of the meaning and function of the Rorschach movement response. However, a post hoc inspection of the data resulted in the subjective impression that an additional and unpredicted change had taken place in the records of the dream-deprived group which might further differentiate it from the non-dream-deprived group: use of space seemed to have increased markedly. To test this impression insofar as it was possible to do so under the circumstances, all 100 records were also scored for space percepts by the investigator. Separate tabulations were made for primary and secondary space (PS and SS), and the differences between each subject's test and retest space scores (PS-D and SS-D) were used as the measures of change in each case. However, since the investigator's scoring was not completely blind, her results were compared with those of a third scorer who had previously done full-scale blind scoring on 10 cases (20 records). Spearman rank-order correlation produced an r_s of .84, indicating that the space scoring was reasonably reliable and that the investigator's scoring was not biased in favor of the hypothesis.

With this evidence of the objectivity of the space scoring, the hypothesis that use of space increases following dream deprivation was tested by comparing the experimental and control groups on the basis of Mann-Whitney *U* tests. As can be seen from Table 5, results for primary space were highly significant, while results for secondary space were nonsignificant in experimental versus placebo-control comparisons and less significant in experimental versus drug-control comparisons.

These results seem to support Fonda's (1960) views on the differential significance of primary as opposed to secondary space responses. His interpretation of the meaning of the differential significance of primary space also seems apt:

> the rate of S-emission seems not to be highly correlated with the actual pressure of *external* events. It follows, therefore, that forces conceived as having power to determine S-emission rate, though they be experienced as arising in reaction to heteronomous events, actually emanate from within, from the *unconscious* regions of the personality, and so might properly be regarded as "ego-alien" [pp. 89–90].

In this study, subjects certainly did not seem more resistant to the examiner, the task, or any other external factor on retest. Instead, they seemed to be trying to resist or at least control an internal pressure, a threatened engulfment by their own fantasies as shown by the increased number and increasingly disorganized and disturbing character of their movement reponses. The fact that such an increase in movement existed side by side with an increase in space is no contradiction to this interpretation, for, as Fonda (1960) notes: "Emphasis upon S in a record certainly gives no guarantee that the struggle for autonomy or mastery [over ego-alien impulses] has been successful. Actually, of course, S per cent is only a measure of effort, not of achievement [p. 100]." Thus, the heightened space scores in the dream-deprived subjects reflect their struggle to resist an intense upsurge of kinesthetic fantasy, and their increased movement scores reflect the failure of that effort in most cases. In some instances, this failure is more dramatically illustrated: nine experimental subjects used movement as a determinant for primary space responses on retest, while only four controls did so. In sum, for the purposes of this study, Fonda's relatively specific view of movement might be narrowed down still further as in the following paraphrase: space is resistance to movement.[12]

---

[12] If this interpretation is correct, then it is conceivable that the presence of primary space in a record may be a reliable sign of capacity for movement even when no movement responses occur.

TABLE 6

MEAN SCORES ON FOUR VARIABLES FOR FOUR CATEGORIES OF SUBJECTS

Variable	Experimental subjects		Placebo subjects	
	True positives	False negatives	False positives	True negatives
$\bar{x}$ Initial WTM	33.6	27.2	34.0	38.4
$\bar{x}$ WTM-D	14.0	−1.4	13.6	−1.6
$\bar{x}$ PS-D	2.6	2.0	0.2	−0.6
$\bar{x}$ WMD-D	1.2	2.0	0	0.2
Combined $\bar{x}$ PS-D and WMD-D	3.8	4.0	0.2	−0.4

On the basis of the foregoing findings with regard to both movement dissolution and space, it seemed worthwhile to take another look at the data on quantity of overall movement in relation to dream deprivation. Although quantity of movement increased significantly following dream deprivation while remaining approximately stable following non-dream-deprivation, not all dream-deprived subjects showed a striking rise in movement, nor did all control subjects show a stable or decreased amount of movement.

The experimental and placebo-control groups were the best matched in terms of initial (preexperimental manipulation) quantity of movement as measured by WTM-D. The combined median for these two groups was 3.5. Using this as a cutting score to divide these 40 subjects into those with high and low movement difference scores (WTM-D) in each group (experimental versus control), there were five false negatives (experimental lows) and five false positives (control highs). To examine the possibility that these two "false" groups might still be differentiable from their respective counterparts, true positives and true negatives, each "false" case was matched with the "true" case whose WTM-D score most closely approximated his own. This resulted in four groups of five subjects each. Success of the matching as well as direction of the differences can be judged from Table 6, which shows mean scores for each group on four variables: initial WTM, WTM-D, PS-D, and WMD-D, and for the latter two combined. It can be seen that the matching was very close indeed and that differences were in the predicted direction. In addition, a Mann-Whitney *U* test on combined PS-D and WMD-D scores for these 20 subjects is highly significant ($p = > .02$). Thus, even the least reactive group of experimental subjects is distinguishable from its control counterpart, and, conversely, even the most reactive control is distinguishable from its experimental counterpart.

A simpler way of stating the above, at least in part, is to note that only 2 of 20 experimental subjects failed to show a post-dream-deprivation rise in at least one of three variables: overall quantity of movement, amount of movement dissolution, or primary space. To be sure, those who did react to one, two, or all three of these variables did so to varying degrees, and the question of what determines threshold and intensity or degree of response is an interesting one. Before indulging in such speculations, however, it would appear wise to first make sure (*a*) that the overall results hold up under cross-validation, at least with a comparable high M sample, and (*b*) that the two nonreactors in the present study did not simply escape through a loophole in the methodology, since this study contained no control on whether or not subjects actually took the dream-depriving drugs as they had agreed to do. (This last loophole could be closed by having subjects sleep in the lab and using awakening rather than drug-induced dream deprivation.)

At any rate, the overall results of this study might best be highlighted by quoting Bendick and Klopfer's (1964) comment on a very different but highly relevant study of their own:

> This [study] seems to lend weight to those theorists who had regarded the process of projecting movement into the blots in terms of kinesthetic symbolism in a rather concrete sense. It may be rather surprising to those other theorists who have thought of movement responses as being reflections of inner activity of an ideational sort which are then reflected back from the Rorschach cards after having attributed to them for the purpose of projection [p. 263].

Finally, it should be noted that this study also lends weight to those theorists (e.g., Beck, 1961, p. 72) who have regarded the similarity between movement responses and dreams in a rather concrete sense. It, too, may be rather surprising to those other theorists who have thought of both movement responses and dreams in a more general, abstract, visual-ideational sense.

REFERENCES

Aserinsky, E., & Kleitman, N. Regularly occurring periods of eye motility and concomitant phenomena during sleep. *Science,* 1953, **118**, 273–274.

Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *Journal of Applied Physiology,* 1955, **8**, 1–10.

Auble, D. Extended tables for the Mann-Whitney Statistic. *Bulletin of the Institute of Educational Research at Indiana University,* 1953, **1**(2), iii, 1–39.

Beck, S. J., Beck, A. G., Levitt, E. E., & Molish, H. B. *Rorschach's Test I. Basic Processes.* (3rd ed.) New York: Grune & Stratton, 1961.

Bendick, M. R., & Klopfer, W. G. The effect of sensory deprivation and motor inhibition on Rorschach movement responses. *Journal of Projective Techniques,* 1964, **28**, 261–264.

Cronbach, L. J. Statistical methods applied to Rorschach scores: A review. *Psychological Bulletin,* 1949, **46**, 393–429.

Dement, W. The effect of dream-deprivation. *Science,* 1960, **131**, 1705–1707.

Dement, W., Clemes, S., Keating, G., & Ruster, G. Further studies of the effect of dream-deprivation. Paper presented at the Fourth Annual APSS meeting, March, 1964.

Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1957, **9**, 673–690. (a)

Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346. (b)

Dement, W., & Wolpert, E. A. The relation of eye movements, body motility and external stimuli to dream content. *Journal of Experimental Psychology,* 1958, **55**, 543–553.

Fisher, C., & Dement, W. Studies on the psychopathology of sleep and dreams. *American Journal of Psychiatry,* 1963, **119**, 1160–1168.

Fiske, D. W., & Baughman, E. E. Relationships between Rorschach scoring categories and the total number of responses. *Journal of Abnormal and Social Psychology,* 1953, **48**, 25–32.

Fonda, C. P. The white-space response. In M. A. Rickers-Ovsiankina (Ed.), *Rorschach psychology.* New York: Wiley, 1960.

Goldman, A. E., & Herman, J. L. Studies in vicariousness: The effect of immobilization on Rorschach movement responses. *Journal of Projective Techniques,* 1961, **25**, 164–165.

Holtzman, W. H., Thorpe, J. S., Swartz, J. D., & Herron, E. W. *Inkblot perception and personality.* Austin: University of Texas Press, Hogg Foundation Research Series, 1961.

Klopfer, B., Ainsworth, M. D., Klopfer, W. G., & Holt, R. R. *Developments in the Rorschach technique.* New York: Harcourt, Brace & World, 1954.

Klopfer, B., & Kelly, D. M. *The Rorschach technique.* New York: World Book, 1942.

Loveland, N. T., & Singer, M. T. Projective test assessment of the effects of sleep deprivation. *Journal of Projective Techniques,* 1959, **23**, 323–334.

Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: A test of the sensory-tonic theory. *Journal of Personality,* 1953, **21**, 400–410.

Mourly Vold, J. *Ueber den traum. Experimental-psychologische untersuchungen.* O. Klemm (Ed.), 2 vol. Leipzig: Barth, 1910–1912.

Neel, A. F. Inhibition and perception of movement on the Rorschach. *Journal of Consulting Psychology,* 1960, **24**, 224–230.

Oswald, I., Berger, R., Jaramillo, R., Keddie, K., Olley, P., & Plunkett, G. Melancholia and barbiturates: A controlled EEG, body and eye movement study of sleep. *British Journal of Psychiatry,* 1963, **109**, 66–78.

Palmer, J. O. Alterations in Rorschach's experience balance under conditions of food and sleep deprivation: A construct validation study. *Journal of Projective Techniques,* 1963, **27**, 208–213.

Rechtschaffen, A., & Maron, L. The effect of amphetamine on the sleep cycle. *Electroencephalography and Clinical Neurophysiology,* 1964, **16**, 438–445.

Rorschach, H. *Psychodiagnostics: A diagnostic test based on perception.* (1st ed., 1921) Bern: Hans Huber, 1951.

Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach human movement responses. *Journal of Consulting Psychology,* 1954, **18**, 325–331.

Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *Journal of Consulting Psychology,* 1952, **16**, 359–364.

Singer, J. L., & Opler, M. K. Contrasting patterns of fantasy and motility in Irish and Italian schizophrenics. *Journal of Abnormal and Social Psychology,* 1956, **53**, 42–47.

Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience type. *Journal of Consulting Psychology,* 1954, **18**, 1–9.

Voigt, W. H., & Dana, R. H. Inter- and intrascorer Rorschach reliability. *Journal of Projective Techniques,* 1964, **28**, 92–95.

Williams, H. L., Hammack, J. T., Daly, R. L., Dement, W. C., & Lubin, A. Responses to auditory stimulation, sleep loss and the EEG stages of sleep. *Electroencephalography and Clinical Neurophysiology,* 1964, **16**, 269–279.

(Early publication received May 12, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 87–94

THE REDUCTION OF UNCERTAINTY CONCERNING FUTURE PAIN[1]

AUSTIN JONES
University of Pittsburgh

P. M. BENTLER
University of California at Los Angeles

GEORGIA PETRY
Psychological Service of Pittsburgh

Ss received series of electric shocks to the forearm in which the temporal schedule and the sequence of shock intensities were associated with varying levels of randomness or uncertainty. The Ss were permitted an instrumental response which produced, in advance, information concerning 1 or both of these aspects of the shocks. Such information functioned as a strong positive reinforcement for most Ss. Instrumental response rate was a significant increasing function of the degree of uncertainty associated with the shock series, and occurred significantly more often for information concerning the temporal occurrence of shock than for its intensity.

Studies of information deprivation have shown that the reduction of uncertainty regarding series of photic stimuli reinforces instrumental response and that the rate of response is an increasing linear function of the magnitude of the uncertainty reduced (Jones, 1961, 1964, 1966; Jones, Wilkinson, & Braden, 1961). In these experiments the reduction of uncertainty through the presentation of series of brief light flashes varying in their statistical properties appears to constitute a reinforcement principle that is independent of any observable, direct association with a previously established primary or secondary reinforcement. This suggests that the reduction of uncertainty may be a dimension of reinforcement of considerable generality—serving to modify a wide range of human behaviors.

The purpose of the present study was to assess the human motivational processes associated with the reduction of uncertainty concerning future pain. The subjects experienced various conditions with the instruction that some brief, painful electric shocks to the forearm would occur, but that the schedule of shocks and their intensity would be probabilistically determined, so that the subject would always be in some measure uncertain as to the characteristics of the stimulation he would receive. The subject was permitted an instrumental response which would provide information (uncertainty-reduction) concerning future shocks in the form of messages transmitted via earphones. In the first of two experiments, it was hypothesized (*a*) that the majority of subjects would perform such responses, (*b*) that response rate would be an increasing function of the degree of uncertainty associated with the temporal sequence of the shocks, and (*c*) that response rate would be greater when associated with the reduction of uncertainty concerning two stimulus dimensions rather than one. The expectation that the majority of subjects would perform the uncertainty-reducing responses was based upon the view that the reduction of uncertainty permits the subjects to make responses which serve to minimize pain and/or anxiety. Such responses may be of a "central" or ideational type, such as thinking to oneself "I don't have to be anxious about the shock now because it is not due for 1½ minutes," or of a motor type, such as waiting relaxed till the shock

---

[1] This study was supported in part by Grant MH 07632 from the National Institute of Mental Health, United State Public Health Service. Experiment II is based upon a master's thesis submitted by the third author to the Graduate Faculty of the University of Pittsburgh.

is due and then gritting the teeth or making other postural adjustments which provide kinesthetic feedback serving to "compete" with painful stimulation. Messages which reduce uncertainty would then be expected to function as strong positive secondary reinforcements because of their role in mediating the reduction of pain and/or anxiety. The principle of secondary reinforcement may be applied in a more direct fashion to the present problem, however, to yield an opposing hypothesis. As the uncertainty-reducing messages are always followed by shock, the messages must be presumed to acquire some negative secondary reinforcement value, thus leading to a tendency not to respond. Some animal experimentation indicates that a warning signal may even become more aversive than the shock with which it is associated (Sidman & Boren, 1957). In the present experiment the potential positive and negative secondary reinforcement factors are not mutually exclusive, and the experimental design permits an assessment of their relative strength.

EXPERIMENT I

Method

Subjects. Thirty-two male undergraduates served as paid volunteer subjects. They understood prior to volunteering that electric shocks were involved and that they would be permitted individually to designate an upper limit of shock intensity which would be observed throughout the experiment.

Apparatus. Shocks of 1.5-second duration were administered through silver electrodes to the subject's nonpreferred forearm. Shock intensity was controlled by a Variac. The subject sat before a panel containing a well-illuminated clock dial placed at eye level. The experimenter sat in an adjacent room before a second clock synchronized with the first. The subject held in his preferred hand a button with which he could signal a request for information to the experimenter, and wore earphones through which the messages were received.

Experimental design. Subjects were randomly assigned to two experimental groups. Group 1 ($N = 18$) was permitted to respond for information (uncertainty-reduction) concerning the temporal occurrence of the shocks. Group 2 ($N = 14$) was permitted to respond for information concerning both the temporal occurrence and the intensity of the shocks. Each subject received 12 shocks in each of three consecutive 12-minute conditions separated by 3-minute rest periods. The three conditions corresponded to low, medium, and high degrees of temporal uncertainty of the shocks. The subjects understood that the shocks could occur only at the quadrants of the clock dial, that is, as the second hand crossed 3, 6, 9, or 12. Under low uncertainty, the 12 shocks were scheduled randomly within the 12-minute period but with the constraint of a 45-second minimum interval between shocks. Under medium uncertainty, the 12 shocks were randomly scheduled with a minimum between-shock interval of 30 seconds; and under high uncertainty the between-shock constraint was reduced to 15 seconds (technically no constraint at all, as the time scale employed in each of the conditions had 15 seconds as its "unit" value). The three conditions were administered to the subjects in counterbalanced order.

The intensity of the shocks was in each condition varied randomly over five levels which were defined with reference to each subject's preliminary scaling of shock tolerance. The subject's minimum-perceptible and maximum-tolerable shocks were determined in Variac units. Intensity Level 2 was set at the mid-point of that range and Level 5 at the top (i.e., maximum-tolerable), with Levels 3 and 4 representing equal Variac intervals between them. Level 1 was determined by subtracting one equal Variac interval from the mid-point of the range. The shock levels determined in this fashion appeared, collectively, to be substantially aversive for each subject, yet readily discriminable. In each of the three conditions, the sequence of shock levels was randomized with the constraint that no more than two shocks of the same level may occur consecutively.

Procedure. The subjects were run individually. In printed instructions, the subject read an explanation of the preliminary shock scaling and of the first of the three conditions. The instructions emphasized that the subject was free to respond as often or as seldom as he wished, that the shocks were scheduled automatically by the apparatus and totally unrelated to subject's responses, and that the experiment was not concerned in any way with possible personality implications of the subject's pattern of responses. The subjects understood that response for information was permitted for only one shock in advance; that is, subjects were not free to accumulate information about two or more shocks by rapidly repeated button presses.

The shock scaling was carried out by administering an ascending series of shocks in small increments, starting from zero. Following determination of the subject's minimum-perceptible and maximum-tolerable shocks, the five levels were computed. The subject was informed of the principle of randomization over the five levels and was given a sample shock of each. During the experiment the shocks were administered manually by the experimenter. In order, however, to avoid the social-psychological implications of punishment personally administered, the subject was told that the shocks were determined automatically by the apparatus in accordance with predetermined schedules. The experimenter then explained the use of the clock. The subject was to note the position of the second hand whenever he would press his button to indicate a request for information. The experimenter would then announce over the earphones the time remaining before shock onset. This message was always given in 15-second units (e.g., "45 seconds," "1 minute, 15 seconds"). The subject was instructed to begin the count of that interval from the quadrant of the dial immediately preceding the position of the second hand at the time of the button press. For example, if the subject pressed his button when the second hand was at some point between 3 and 6, and if the message was "45 seconds," he would know that the next shock would occur when the second hand next passed 12. The subject was rehearsed briefly to ascertain his ability to interpret the messages correctly. The procedure described thus far applied to the subjects of Group 1. Group 2 was treated identically except that each button press was followed by a message indicating the level of the shock as well as occurrence (e.g., "1 minute 30 seconds—level five").

Prior to beginning the first condition, the minimum interval between shocks was explained to the subject, but without reference to the constraints applying to the remaining conditions. The constraints for the second and third conditions were explained to the subject over the earphones immediately prior to their beginning. From the beginning of the first condition to the end of the third, the experimenter remained out of sight in an adjacent room. At the conclusion of the experiment a brief questionnaire was administered which was concerned with the aversiveness of the procedure, the subject's rationale for his pattern of response or lack of it, and other points intended as informal guidance for the experimenters.

Results

Eight of the 32 subjects made no responses, five in Group 1 and three in Group 2. The number of responses made by the remaining 24 subjects ranged from 2 to 36, with an overall mean of 16.75. (The mean response rates of Groups 1 and 2, including data from nonresponders, were 11.06 and 14.50 respectively, and do not differ significantly; $t < 1.0$.)

In Figure 1, proportional response rate is plotted as a function of the degree of uncertainty associated with the three conditions. The data consisted of each subject's response rate for each condition expressed as a proportion of their total. Thus for each subject three proportions summing to one were computed. The proportional transformation was employed in order to remove variance associated with individual differences in absolute response rate, which are irrelevant for an assessment of this function. The analysis of variance of trends (Grant, 1956) shows response rate, pooled over the two groups, to be a significant linear function of degree of information ($F(1,23) = 8.76$, $p < .01$). Considered separately, the functions for Group 1 and Group 2 are each significant—F lin. $(1,12) = 4.88$, $p < .05$, and overall F $(1,10) = 4.38$, $p < .05$, respectively. The differences between the functions are not significant ($F < 1.0$).

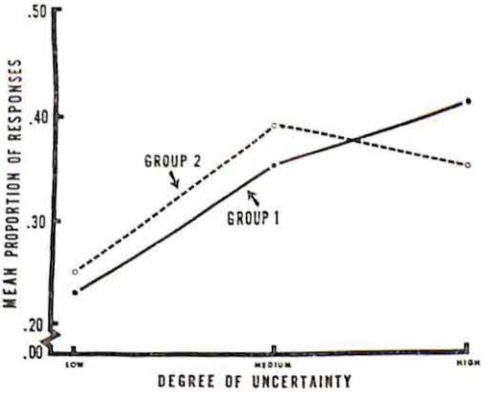


FIG. 1. Experiment I: Mean proportion of responses as a function of degree of uncertainty.

Discussion

The results support the view that the reduction of uncertainty concerning future pain constitutes a positive reinforcement for the majority of the subjects. Their answers to the postexperimental questionnaire suggested that the motivation to respond rose sharply when an unusual interval had passed without shock, that is, when in a state of high temporal uncertainty, and whenever a shock of high intensity was anticipated.

The hypothesis that response rate would be an increasing function of the degree of temporal uncertainty was supported by analysis of the data for Group 1 and for Groups 1 and 2 pooled (see Figure 1). The analysis for Group 2 showed significant differences in response rate between the uncertainty levels but failed to show either the linear or quadratic components as significant ($.05 < p < .10$ for both). As the observed downward inflection at high uncertainty for Group 2 is not reliable, it would appear that the pooled function, which is significantly linear at

$p < .01$, may be regarded as the best estimate of the true relationship of response rate to level of temporal uncertainty. The increasing trend of responses over uncertainty values is consistent with the theory that higher degrees of uncertainty are associated with parallel increments in anxiety, and that it is anxiety which motivates the uncertainty-reducing responses.

Support for the third hypothesis, that response rate would be greater when associated with the reduction of uncertainty concerning two stimulus dimensions rather than one, was not obtained. Subjects receiving information about the schedule and intensity of shocks did not respond significantly more often than those receiving information only about the schedule. This would seem to suggest that uncertainty-reduction concerning the temporal occurrence of shock is of greater motivational relevance than uncertainty-reduction concerning shock intensity, possibly because of its greater role in the control of anxiety. In this context, certain deficiencies of the experimental design become apparent insofar as a test of the hypothesis relating response rate to number of dimensions of uncertainty-reduction is concerned. If it should be true that temporal uncertainty is of greater motivational relevance, then an experiment in which Groups 1 and 2 respond for intensity information and for intensity-plus-temporal information, respectively, would be expected to have results opposite to those of the present experiment. That is, the use of temporal information as the additional dimension of information in Group 2 should result in a greatly and significantly increased response rate. Thus the present experimental design could not be regarded as adequate in this regard unless the assumption is made that the different dimensions of uncertainty are of equal motivational relevance. A further limitation of the design is suggested by the great individual differences in absolute response rate. As the comparison in question is based upon the data of independent subject groups, and as the obtained difference was in the hypothesized direction, the problem of Type 2 errors is raised. The magnitude of the individual differences in absolute response rate suggests that a repeated-measures design would have been more appropriate for a test of the hypothesis.

EXPERIMENT II

The purpose of the second experiment was to provide (*a*) a more adequate test of the hypothesis that response rate will be greater for two dimensions of uncertainty-reduction than for one, (*b*) a further assessment of the shape of the function relating response rate to the degree of uncertainty within stimulus dimensions, and (*c*) a comparison of the motivational relevance of uncertainty-reduction regarding the intensity of shocks and their temporal sequence. A further intent of the experiment was the development of a technique which would elicit a higher level of motivation for uncertainty-reduction than that occurring in Experiment I. The general approach was to increase the absolute uncertainty values of the stimulus dimensions and to sensitize subjects to them by the instructional set.

Method

Subjects. Twenty-four male undergraduates served as paid volunteer subjects. As in Experiment I, subjects understood in advance that electric shock was involved and that they would be permitted to designate the upper limit of shock intensity. Each subject was required to meet a shock-intensity criterion and a response-rate criterion in order to be retained in the study. The intensity criterion was the establishment of an upper limit at least 35 Variac dial units above the minimum-perceptible intensity, in order to assure that seven shock levels would be clearly discriminable. In order to increase the reliability of the proportional response rates which were employed heavily in the analysis of data, a minimum of 10 responses over the total experiment was required.

Apparatus. The subject sat before an illuminated clock dial below which two response buttons were mounted on a wooden panel—one labeled "time," the other "intensity." A continuous record of response was kept by an Esterline-Angus recorder. Shock duration was 2.0 seconds. The apparatus was in all other respects similar to that used in Experiment I.

Experimental design. A repeated measures design was employed in which all subjects received three conditions corresponding to low, medium, and high uncertainty, within each of which response for temporal information, intensity information, or both was permitted. The level of uncertainty characterizing a particular condition was the same for the temporal and intensity dimensions.

The levels of uncertainty were operationally defined by the degree to which the intensity and the intershock interval pertaining to each shock varied randomly from those associated with the immediately preceding shock. There was a maximum of seven "variation units" for both the temporal and intensity dimensions. For the intensity dimension the seven units corresponded to seven equally spaced Variac values ranging from Level 1, which was set at five Variac dial units above the minimum-perceptible shock intensity, to Level 7, which was the subject's upper limit of intensity. The units of the temporal dimension consisted of the seven intershock intervals ranging from 15 seconds (Level 1) to 1 minute 45 seconds (Level 7) in 15-second increments.

Under the low uncertainty condition, the interval and intensity of each shock were varied randomly among the three values closest to the interval and intensity associated with the preceding shock. Thus if a particular shock was at Level 4 intensity and had occurred after a Level 5 interval (1 minute 15 seconds), the intensity of the next shock would be randomly drawn from Levels 3, 4, and 5, and the intershock interval would be randomly drawn from Levels 4, 5, and 6 (1 minute; 1 minute 15 seconds; and 1 minute 30 seconds).

Under medium uncertainty, the intensity and interval of a particular shock were randomly selected from among the five levels closest to the values associated with the preceding shock. Under high uncertainty the intervals and intensities were drawn randomly from the total range of seven levels.

The duration of each condition was 14 minutes, within which 14 shocks were administered—two at each of the seven intensity levels. The initial shock in each condition was randomly selected from the seven levels.

Procedure. Instructions similar to those of Experiment I were given to the subjects orally. They were informed that the "time" and "intensity" buttons could be pressed singly, in either order, or simultaneously. The only restrictions on response were that none should be made while receiving a shock, and that use of a particular button should not be repeated until an intervening shock had occurred. It was again emphasized that subjects were to feel free to respond as often or as seldom as they wished. At the beginning of each condition, a 3 × 5 inch card indicating the set of intershock intervals in effect and the rule governing the determination of shock intensities was placed under the clock dial before which the subject sat. The procedure was in all other ways the same as that employed in Experiment I.

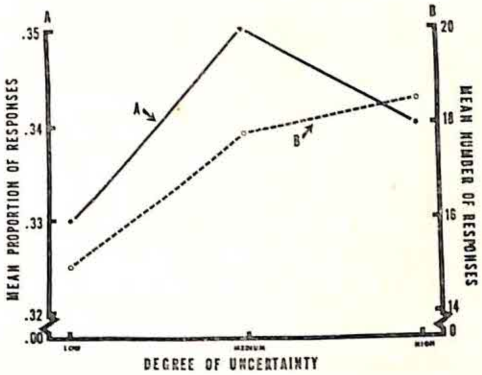


FIG. 2. Experiment II: Response as a function of degree of uncertainty, plotted (A) proportionally, and (B) as absolute frequencies. (Responses for temporal and intensity formation are pooled.)

Results

The absolute response rate was substantially higher than in Experiment I. Only three subjects were excluded for failure to meet the criterion of 10 responses. The total number of trials was 42. For the 24 subjects meeting the criterion, the number of trials on which one or both buttons were pressed ranged from 10 to 42 with a mean of 34.75.

The hypothesis that response rate would be greater for two dimensions of uncertainty-reduction than for one was assessed by comparing the proportion of the trials on which both buttons were pressed with the proportion on which only one button was pressed. The difference was in the expected direction but was not significant ($t(23) = 1.61$, $p > .05$).

In Figure 2A proportional response rate is plotted as a function of degree of uncertainty. The analysis of variance of trends shows the overall function to be significant ($F = 3.44$, $df = 2/46$, $p < .05$), but with neither the linear nor quadratic trends attaining significance ($F = 3.17$ and 3.50, respectively, with $df = 1/23$ and $p > .05$). In the interest of gaining further information concerning the probable shape of the "true" function relating response rate to degree of information, a similar analysis was carried out upon the absolute response rate data (Figure 2B). This analysis of the function showed the linear component, but not the quadratic, to be significant—F lin. $(1/23) = 4.91$, $p < .05$, and F quad. $(1/23) = 1.36$, $p > .05$.

Speed of response was also analyzed as a function of degree of uncertainty, with individual subjects' speed scores computed as the mean of the reciprocal latencies of the responses. Data for a particular subject were excluded whenever there was an absence of response at any of the three degree-of-uncer-

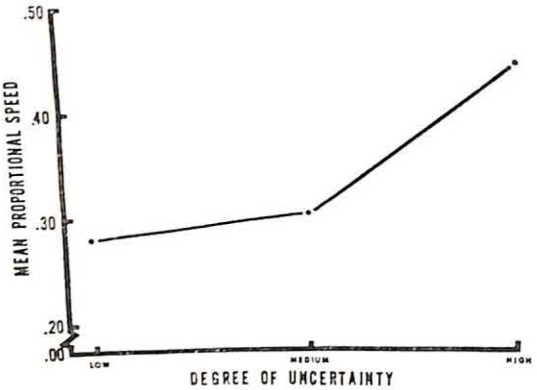


Fig. 3. Experiment II: Mean speed scores expressed proportionally as a function of degree of uncertainty. (Data from trials in which responses for both temporal and intensity formation were made.)

tainty conditions, in order that absence of response would not be confounded with zero latency; thus the *N*s for these analyses are less than 24. The nature of the speed function was found to be related to the number of responses made per trial. Considering the trials in which responses for both temporal and intensity information were made ("both" trials hereafter), response speed was shown to be a significant linear function of degree of uncertainty—(see Figure 3) F lin.$(1,16) = 13.60$, $p < .01$, and F quad. $(1,16) = 1.50$, $p > .05$. Considering the trials in which only one response was made ("single" trials hereafter), neither the overall trend nor either component was significant—overall F and F lin. < 1.0, F quad. $(1/16) = 1.20$, $p > .05$. The speed functions for the "both" trials and for the "single" trials differ significantly as tested by the linear component of the interaction (F $(1/9) = 9.40$, $p < .02$). The mean absolute speed scores, however, pooled over uncertainty conditions, were significantly higher (faster) for the "single" trials than for the "both" trials (F $(1/9) = 11.64$, $p < .01$).

The motivational relevance of the temporal and intensity dimensions of uncertainty-reduction was assessed by a comparison of absolute response rates for the two dimensions. Response rate for temporal information was significantly greater than for intensity information ($t = 2.19$, $df = 23$, $p < .05$). The means were 29.0 and 22.8, respectively. Considering "single" trials only, the effect was magnified, the two means being 5.7 and 12.7. The absolute speed scores, however, were significantly higher for responses for intensity information than for temporal information ($t = 2.16$, $df = 21$, $p < .05$).

Speed functions over degrees of uncertainty were analyzed separately for responses for temporal and intensity information. For intensity information, response speed was an increasing linear function of degree of uncertainty (F lin. $(1/19) = 10.13$, $p < .01$). For temporal information, response speed was not significantly related to degree of uncertainty—overall F $(2/42) = 3.05$, F lin. $(1/21) = 3.85$, F quad. $(1/21) = 1.88$, each with $p > .05$.

Discussion

The results of Experiment II provide strong additional support for the view that the reduction of uncertainty concerning future pain functions as a positive secondary reinforcement for the majority of subjects. Eighty-eight percent of the subjects met the criterion of 10 responses. The average percentage of total trials on which one or both of the possible responses occurred was approximately 83%. Thus the positive reinforcement potential associated with the uncertainty-reducing messages appears to have substantially outweighed the negative secondary reinforcement potential which must be assumed to have developed to some degree because the messages were regularly followed by shock. The exact mechanism or mechanisms involved in the reinforcement were not considered directly in this study. Questionnaire data and informal discussion with subjects, however, appear to support the formulation that uncertainty concerning future pain elicits anxiety and that the uncertainty-reducing responses mediate the reduction of anxiety through subsequent ideational and postural-muscular responses. It appears further that the ideational responses, for example those by which the subject reassures himself that he will not be shocked until a particular period has elapsed, are somewhat more prominent in the control of anxiety than the postural-muscular responses which may reduce pain magnitude by providing "competing" sensory input.

If the reduction of uncertainty is positively reinforcing, then the rate of instrumental response should be an increasing function of the amount or degree of uncertainty which is reduced. The analysis of variance of trends applied to proportional response rates showed significant differences associated with degree of uncertainty, but with neither the linear nor quadratic components attaining significance (Figure 2A). A parallel analysis of the absolute response rates (Figure 2B) showed a significant increasing linear function. Proportional treatment of response rate is ordinarily preferable for such repeated measures designs in which individual differences in absolute response rate are irrelevant, but the alternative analysis is useful in view of the ambiguity associated with the first. The combined results of Experiments I and II suggest that the function relating response rate to degree of uncertainty is monotonic and negatively accelerating.

Experiment II provided a second and independent measure of response strength—speed. Considering the trials in which responses for both temporal and intensity information occurred, proportional response speed was shown to be a significant increasing linear function of degree of uncertainty (Figure 3), thereby broadening the support for the formulation that uncertainty-reduction regarding pain constitutes a positive reinforcement. For trials in which a single response was made, however, no significant trends over degree of uncertainty were found. A possible explanation for this discrepancy is suggested by the finding that the absolute speed scores were significantly higher for the "single" trials than for the "both" trials. It appears likely that the relatively prompt response on "single" trials may have substantially attenuated anxiety, and that the "both" trials, in which response was slower, may have been associated with substantially higher levels of anxiety. The failure of discrimination of uncertainty levels as measured by response speed on "single" trials may reflect simply a much lower level of anxious arousal.

The hypothesis that response rate would be significantly greater for two dimensions of information than for one was not supported in Experiment I. In the present experiment, a more sensitive test of this hypothesis was provided by means of a repeated-measures comparison of the number of trials on which both responses were made and the number on which a single response was made. The difference was again nonsignificant. Comparison of the proportional response rates for temporal and intensity information, however, provides an interpretation of this negative result. Response rate for temporal information was significantly greater than for intensity information, and there was but a very slight and nonsignificant difference in the number of trials for which both responses were made and the number for which only temporal responses were made. Thus it appears that temporal information was very much more motivationally relevant than intensity information, and that the additional intensity information occurring on the "both" trials constituted but a minor increment in reinforcement. This interpretation is consistent also with the findings of Experiment I, in which subjects allowed to respond for temporal or for temporal plus intensity information showed a but slightly and nonsignificantly higher response rate for the latter. Due to the differential motivational relevance of the two dimensions employed in the present study, the hypothesis that response rate is a function of the number of dimensions of uncertainty-reduction remains essentially untested and awaits further experimentation in which the motivational equivalence of the dimensions employed is established in preliminary scaling.

Evaluation of the relative reinforcement values of temporal and intensity information is complicated by the finding that, while the frequency measure of response strength shows temporal information to be superordinate, the speed of response measure was significantly greater for intensity information and was significantly related to degree of uncertainty for intensity information only. A clear explanation of this seeming paradox does not appear to be provided by the present data.

REFERENCES

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychological Bulletin*, 1956, **53**, 141–154.

JONES, A. Supplementary Report: Information deprivation and irrelevant drive as determiners of an instrumental response. *Journal of Experimental Psychology*, 1961, **62**, 310–311.

JONES, A. Drive and incentive variables associated with the statistical properties of sequences of stimuli. *Journal of Experimental Psychology*, 1964, **67**, 423–431.

JONES, A. Information deprivation in humans. In B. Maher (Ed.), *Explorations in personality research*, Vol. III. New York: Academic Press, 1966.

JONES, A., WILKINSON, H. J., & BRADEN, I. Information deprivation as a motivational variable. *Journal of Experimental Psychology*, 1961, **62**, 126–137.

SIDMAN, M., & BOREN, J. L. The relative aversiveness of warning signal and shock in an avoidance situation. *Journal of Abnormal and Social Psychology*, 1957, **55**, 339–344.

(Received February 15, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 95–107

TOWARD A THEORY OF "HYPNOTIC" BEHAVIOR:

EXPERIMENTAL ANALYSES OF SUGGESTED AMNESIA[1]

THEODORE XENOPHON BARBER AND DAVID SMITH CALVERLEY

Medfield Foundation, Harding, Massachusetts

Suggestions of amnesia were given to 144 Ss in a $2 \times 2 \times 4$ factorial experiment. The independent variables were: (1) spoken versus recorded presentation of suggestions; (2) presence versus absence of hypnotic induction; and (3) 4 types of suggestions for amnesia—authoritative ("You will not remember!"), permissive ("Try to forget"), suggestions to simulate, and no suggestions. Variable 1 did not produce important differences in amnesic performance. Variable 2 significantly affected 1 of the 5 tests for amnesia, with the presence of hypnotic induction resulting in less amnesia. Variable 3 exerted the most powerful effect. Irrespective of the presence or absence of hypnotic induction, Ss differed on the tests for amnesia depending on whether they were given suggestions to simulate, authoritative or permissive, or no suggestions.

Williamsen, Johnson, and Eriksen (1965) recently demonstrated that hypnotic subjects who have received suggestions of amnesia show (*a*) more amnesia than waking controls and (*b*) less amnesia than waking subjects who have been instructed to act as if they have forgotten (simulators). Since these findings indicate that hypnotic amnesia is not the same as simulated amnesia and is also not the same as normal recall, they are of critical importance in formulating a theory of hypnotic behavior. Before drawing definitive conclusions from the Williamsen et al. study, however, it is necessary to cross-validate their results. This paper presents several methodological considerations relevant to research in this area and an experiment which replicated and extended the experiment of Williamsen et al.

METHODOLOGICAL CONSIDERATIONS

In the Williamsen et al. experiment all subjects learned six critical words. The hypnotic subjects were then given direct suggestions to forget the critical words, the waking simulators were told to act *as if* they had forgotten, and the waking controls were *not* given suggestions pertaining to amnesia. Since the direct suggestions for amnesia were confounded with the hypnotic treatment and the instructions to simulate amnesia were confounded with a waking treatment, it is not clear to what extent the differences in performance on the tests for amnesia were due to (*a*) the presence or absence of "hypnosis" or (*b*) differences in suggestions or instructions given to the three experimental groups. If a hypnotic group and a waking group had received identical suggestions for amnesia, the two groups may have manifested a comparable degree of forgetting (cf. Barber & Calverley, 1962, 1963b, 1963c). Furthermore, a hypnotic group and a waking group instructed to simulate amnesia may not have differed in performance. In addition, if a hypnotic control group had been employed, it may not have differed from the waking control group in the degree of forgetting. These considerations raise the following question that we attempted to answer in the experiment which is described below: Does a hypnotic group manifest more amnesia than a waking group when both groups are given

---

[1] This research was supported by a grant (MH-07003) from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service.

We are indebted to the following for making participation in the experiment a requirement of their nurses' training courses: Esther Budd and Sarah Frankland of Faulkner Hospital; Eileen White of the Framingham Union Hospital; Alice Webster of Burbank Hospital; and Margaret Taylor of Leominister Hospital. We also wish to thank Charles W. Eriksen for sending to us a prepublication copy of the Williamsen, Johnson, and Eriksen (1965) study.

(*a*) identical suggestions of amnesia, or (*b*) identical instructions to simulate amnesia, or (*c*) neither suggestions of amnesia nor instructions to simulate amnesia?

In light of a series of recent studies, reviewed by Barber (1962a, 1965), a number of additional factors were incorporated into the design of our experiment:

1. There is evidence to indicate that experimenters find it difficult to give identically worded suggestions to hypnotic and waking subjects in the same tone of voice, and the biased tone of voice in which the suggestions are administered affects the subjects' responses (Barber & Calverley, 1964b; Troffer & Tart, 1964). To assess the effects of this variable in our experiment, half of the subjects in each experimental group were given the suggestions by means of a tape recording and the tests for amnesia by means of a written questionnaire. The remaining half were given the suggestions and tests for amnesia orally, as in the Williamsen et al. experiment.

2. Authoritative suggestions of amnesia ("You will not remember!"), as used in the Williamsen et al. experiment, may be less effective with awake subjects than with hypnotic subjects in that such peremptory suggestions may appear incongruous in the waking situation and may be unexpected by the awake subject, whereas they fit into the hypnotic situation and are expected by the hypnotic subject. It appears likely, however, that *permissive* suggestions to "*try* to forget" may not seem incongruous to either the awake or the hypnotic subjects. Consequently, in our experiment, in addition to a waking group and a hypnotic group told authoritatively "You will not remember," a second waking group and a second hypnotic group were told permissively, "Try to forget."

3. In the Williamsen et al. study, none of the subjects were asked, after the experiment, whether they had actually forgotten the critical words which they had not verbalized during the experiment. Instead, Williamsen et al. apparently assumed that (*a*) the hypnotic subjects forgot those words which they did not verbalize and (*b*) the simulators did not forget the critical words, even though they behaved as if they did. These implicit assumptions may or may not be valid. If postexperimental interviews had been conducted, both the hypnotic and the simulating subjects might have stated that they did not forget the critical words during the experiment. Consequently, in the experiment described below, all subjects were given a "Post"-Experimental Subjective Reports questionnaire to ascertain whether their subjective testimony concerning what occurred during the experiment was in harmony with their objective behavior as it had been observed during the experiment.

METHOD

Subjects

The subjects were 144 student nurses (ages 17 to 25) who were required to participate in the experiment. Originally, 148 student nurses were directed by their supervisors to report to the experimental room at a specified time to participate in a "psychological experiment." Of these 148, one did not meet the appointment and three refused to participate when they were told, after they had entered the experimental room, that they were to be hypnotized. The remaining 144 subjects were randomly assigned to 16 experimental groups with 9 to each group.

All students were tested individually. At the end of each test session each was admonished not to discuss the experiment with the other students since the "experiment would be ruined" if subjects knew "what the experiment was about" prior to their participation. Each student stated that she understood the importance of this admonition and promised that she would not discuss the experiment with the other students until all had been tested. We found no evidence that any student broke her promise.

Design

The experiment was designed as a 2 × 2 × 4 factorial. The first independent variable was at two levels: Spoken versus Recorded-Written Presentation (of instructions, suggestions, and tests for amnesia). The second independent variable was also at two levels: Presence versus Absence of Hypnotic Induction. The third independent variable, Types of Suggestions, was at four levels: authoritative suggestions of amnesia, permissive suggestions of amnesia, instructions to simulate amnesia, and no suggestions of amnesia. The 2 × 2 × 4 factorial requires 16 experimental groups as illustrated in Table 1.

Procedure (Independent Variables)

Spoken versus recorded-written presentation. Half of the subjects (Groups 1–8) were given all suggestions and instructions orally as memorized by the experimenter (DSC). The questions comprising the tests for amnesia were also given to these subjects orally and the subjects answered the questions orally.

TABLE 1

EXPERIMENTAL DESIGN: 2 × 2 × 4 FACTORIAL

	Spoken presentation		Recorded-Written presentation	
	Hypnotic induction	No hypnotic induction	Hypnotic induction	No hypnotic induction
Authoritative suggestions of amnesia	Group 1	Group 5	Group 9	Group 13
Permissive suggestions of amnesia	Group 2	Group 6	Group 10	Group 14
Instructions to simulate amnesia	Group 3	Group 7	Group 11	Group 15
No suggestions of amnesia	Group 4	Group 8	Group 12	Group 16

The other half of the subjects (Groups 9–16) were given all suggestions and instructions by means of a tape recording of the experimenter's voice. The questions comprising the tests for amnesia were given to these subjects in written form and the subjects answered the questions by writing. The experimenter thus did not speak to the subjects in Groups 9–16 during the experiment with the exception of four introductory remarks, namely, greeting the subject, instructing her to be seated, and asking her name and age.

Presence versus absence of hypnotic induction. As Table 1 illustrates, half of the subjects were, and half were not exposed to a hypnotic induction procedure. (It should be noted that of those exposed to the hypnotic induction, half were given the induction orally and half by means of a tape recording.)

Subjects exposed to the hypnotic induction (Groups 1–4 and 9–12) were first told that they were to be hypnotized, were asked to fixate on a light blinking in synchrony with the sound of a metronome, and were then given a standardized induction procedure, for a period of 10 minutes, which was adapted from the procedures of Friedlander and Sarbin (1938), Marcuse (1959), and Weitzenhoffer and Hilgard (1959). This hypnotic induction procedure included (*a*) motivational suggestions ("You will find hypnosis a most interesting experience . . . . Your willingness to cooperate and your interest is what I ask for"), (*b*) repeated suggestions of eye-heaviness, eye-closure, relaxation, drowsiness, sleep, and of entering deep hypnosis, and (*c*) statements that the subject would now find it easy to respond to further suggestions and to experience suggested effects.[2] Immediately upon completing the induction procedure, the subjects were assessed on response to the first six test suggestions included in the Barber Suggestibility Scale (BSS; Barber & Calverley, 1963a; Barber & Glass, 1962; Barber, Karacan, & Calverley, 1964). These test suggestions were: arm lowering; arm levitation; hand lock (inability to unclasp one's hands); thirst "hallucination"; verbal inhibition (inability to speak one's name); and body immobility. Immediately after the six test suggestions were administered and scored, the subject was given six critical words to learn as described below.

Subjects not exposed to the hypnotic induction (Groups 5–8 and 13–16) were simply told that they were to be given a test, were asked to listen carefully to the instructions, and then were given the six critical words to learn. In contrast to the subjects in the hypnotic groups, the waking subjects in Groups 5–8 and 13–16 had their eyes open continuously during the experiment.

Presentation of six critical words. The six critical words were presented as follows:

> I want you to listen carefully to the following words. After the whole list is read to you I want you to repeat them to make sure you have learned them. I'll say the words now, after which you will repeat them: BUTTER, DARK, SOFT, KING, SLOW, COLD. Now repeat them. (Pause) I'll say them again: BUTTER, DARK, SOFT, KING, SLOW, COLD. Now repeat them.

Of the 144 subjects, 106 (74%) had learned the six critical words, that is, repeated them correctly, after the second presentation. The 38 subjects who had not learned the words after the second presentation were given a third, and when necessary, a fourth and fifth presentation, until they too repeated them correctly. (Groups 9–16 were given all presentations by a tape recording and Groups 1–8 were given all presentations orally.)

Types of suggestions. After the subject had learned the six critical words, she was given one of four types of suggestions, as follows:

1. Authoritative suggestions of amnesia—Two groups that had received a hypnotic induction (Groups 1 and 9) and two groups that had not received an induction (Groups 5 and 13) were given authoritative suggestions of amnesia for the six critical words, as follows:

> Now I want you to listen carefully to what I say. I just gave you a list of words to learn. You

---

[2] A verbatim account of the hypnotic induction procedure may be obtained without charge from Theodore Xenophon Barber, Medfield Foundation, Harding, Massachusetts, or for a fee from the American Documentation Institute. Order Document No. 8320 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

will not be able to remember what the words were! You will have no memory that you were given any words to learn in this experiment or what the words were. You will forget that you learned any words in this experiment until it is stated that the experiment is over. When it is stated that the experiment is over you will be able to remember everything. Now you will forget each of the words completely and you will forget that you were given any words to learn in this experiment until it is stated that the experiment is over.

2. Permissive suggestions of amnesia—Two groups that had received a hypnotic induction (Groups 2 and 10) and two groups that had not (Groups 6 and 14) were told that it was easy to forget and were instructed permissively to try to forget the six critical words, as follows:

> I'm going to test your ability to control your memory by determining how well you are able to forget something I will tell you to forget. You may have easily forgotten which shoe you put on first this morning or what you read in the newspaper yesterday. We've asked people if they remembered what they read 1 or 2 days ago and the great majority couldn't remember off hand. I often find it very hard to remember what I read in the newspaper or in a novel just a few minutes after I've read it. It really is easy to forget some things.
>
> Now I want you to listen carefully to what I say. I just gave you a list of words to learn. In this next part of the experiment I want you to show me how good you are at controlling your memory by forgetting that you were given any words to learn in this experiment or what the words were. Forget that you learned any words in this experiment until it is stated that the experiment is over. When it is stated that the experiment is over you will be able to remember everything. This can be done. Other subjects that participated in this experiment did very well in this test. They were able to put the words out of their mind and to forget them. Now try to the best of your ability to forget each of the words completely and to forget that you were given any words to learn in this experiment until it is stated that the experiment is over.

3. Instructions to simulate amnesia—Two groups that had received a hypnotic induction (Groups 3 and 11) and two groups that had not received an induction (Groups 7 and 15) were instructed to simulate amnesia, as follows:

> This experiment is designed to measure how well people can simulate or fake posthypnotic amnesia. I want to see how well you can play the role of a hypnotized subject. Now what I want you to do is to act as if you have been deeply hypnotized and given the suggestion that you do not remember that you were given any words to learn in this experiment and that you do not remember any of the words. I want you to act as if you had been told that you would have no memory of having learned any words in this experiment. Act as if you do not remember the words until it is stated that the experiment is over. After it is stated that the experiment is over you will no longer have to act and can tell me the words you just learned. I know you can do a good job. Here is a chance to see how good an actress you are. It's not as hard as you think. Now act as if you have forgotten each of the words. Act as if you have forgotten that you received any words in this experiment. Act as if you have forgotten until it is stated that the experiment is over.

4. No suggestions of amnesia—Two hypnotic groups (Groups 4 and 12) and two waking groups (Groups 8 and 16) were not given any suggestions pertaining to amnesia for the words they had learned.

Assessment of Amnesia (Dependent Variables)

Upon completing the above, the experimenter (DSC) assessed the subject's response to the suggestions for amnesia. Amnesia was assessed by four tests which were originally constructed by Williamsen et al. (Experimental Recall, Recognition of Partial Words, Recognition of Whole Words, and "Post"-Experimental Recall) and by one test devised by us ("Post"-Experimental Subjective Reports).[3] Half of the subjects (Groups 1–8) were given the tests for amnesia orally by the experimenter; the others (Groups 9–16) were given the same tests in written form (on a six-page dittoed questionnaire) as follows

Experimental Recall. Page 1 of the questionnaire stated: "Were you given any words to learn in this experiment? No——. Yes——."

Page 2 stated, "If you did learn some words in this experiment, recall as many of the words as you can and write them below."

If the subject answered "No" on page 1, she was not given page 2.

If the answer to page 1 was "No," the subject received a score of zero on this test. If the answer to page 1 was "Yes," the score was the number of critical words correctly recalled on page 2.

Recognition of Partial Words. Page 3 of the questionnaire stated: "You will be shown words that have parts of their letters missing. You are to try to guess what the words are and write them below." The subject was then presented with 12 words that had parts of their letters missing (modeled after the Street Gestalt figures). These partial words were presented one at a time on 3 × 5 index cards. The maximum time of presentation of each partial word was 30 seconds. The 12 partial words included the 6 critical words and 6 neutral (control) words. As in the Williamsen et al. experiment, the neutral words were: LONG, TABLE, SOUR, WOMAN, WATER, HIGH. Pretesting of several alternate forms resulted in the selection of 12 partial words that evoked the correct response about 25% of the time from a group of pilot subjects. Prior to the beginning of the experiment the order of presentation of the partial words was randomized and this order was held constant for all subjects.

---

[3] Williamsen et al. also used a word association test. Since their experimental groups did not differ on this test, we did not use it in our experiment.

The critical and neutral partial words were scored separately. One point was allotted for each critical word recognized and one point was allotted for each neutral word recognized. Thus the maximum score for recognition of either the critical or neutral words was six points.

Recognition of Whole Words. Page 4 of the questionnaire was worded thus:

> If you had learned any words in this experiment, are there any words below which are words you could have learned? If so, place a check beside the word. SOAP——; BOOK——; SOFT——; PARK——; HIGH——; SLOW——; WATER——; DARK——; QUEEN——; SOUR——; COLD——; LONG——; BUTTER——; WOMAN——; KING——; TABLE——; ROBE——; COOK——.

The words included in this recognition test were the six critical words, the six neutral words that had been used in the preceding Recognition of Partial Words test, and six "dummy" words, listed in random order. One point was allotted for each of the critical words checked.

"Post"-Experimental Recall. Page 5 of the questionnaire stated: "The experiment is over. Now that the experiment is over give as many of the words you were asked to learn at the beginning of this experiment as you can remember. Write them below." One point was given for each of the six critical words correctly recalled.

"Post"-Experimental Subjective Reports. The final page of the questionnaire was given only to those subjects who had manifested some degree of apparent amnesia on the tests of recall, that is, who recalled fewer words on Experimental Recall than on "Post"-Experimental Recall. This page was worded thus:

> When it was stated on the immediately preceding page that "the experiment is over" you recalled —— more words than you did at the beginning of the experiment. Apparently you forgot these —— words. [The preceding blank spaces were filled in by the experimenter.] People forget in various ways. Please check which one of the following applies to you:
>
> A——. I forgot the words in that I did not think about them at all and they did not enter my mind until the end of the experiment.
>
> B——. I forgot the words only partially in that I thought of them at certain times, off and on, during the experiment.
>
> C——. I did not forget the words. I knew them or was thinking of them during the experiment.

TABLE 2

MEAN SCORES OF THREE GROUPS ON EXPERIMENTAL AND "POST"-EXPERIMENTAL RECALL AND RECOGNITION OF PARTIAL AND WHOLE WORDS IN PRESENT INVESTIGATION (WITHOUT PARENTHESES) AND IN WILLIAMSEN ET AL. INVESTIGATION (IN PARENTHESES)

	Hypnotic amnesia	Simulation	Control
(a) Experimental recall			
High susceptible	1.5 (1.3)		
Low susceptible	5.4 (3.8)		
Average	3.7_b (2.6_b)	0.0_c $(0.0_c$	5.7_a (5.4_a)
Post-experimental recall			
High susceptible	5.2 (4.6)		
Low susceptible	5.2 (4.8)		
Average	5.2_a (4.7_a)	5.2_a (4.8_a)	5.4_a (5.5_a)
(b) Partial words			
Critical words	4.1_a (3.4_a)	2.3_b (1.6_b)	3.8_a (4.2_a)
Neutral words	1.8_a (1.1_a)	1.3_a (1.6_a)	0.7_a (1.1_a)
(c) Whole words			
High susceptible	3.0 (2.9)		
Low susceptible	5.4 (5.1)		
Average	4.3_a (4.0_a)	1.6_b (0.0_b)	5.0_a (5.5_a)

Note.—Means in the same row containing different letters in the subscript differ from each other at the .05 level of confidence.

Items A, B, and C received weights of 2, 1, and 0, respectively. Subjective scores were derived by multiplying the number of words the subject had ostensibly forgotten (the difference between the scores on "Post"-Experimental and Experimental Recall) either by 2 (if she checked Item A), or by 1 (if she checked Item B), or by 0 (if she checked Item C). (Subjects who had not manifested amnesia, that is, who gave as many words on Experimental Recall as on "Post"-Experimental Recall, were not given the "Post"-Experimental Subjective Reports questionnaire and were automatically assigned a subjective score of zero.)

RESULTS

Replication of Williamsen et al. Experiment

Three of the 16 groups in our experiment (Groups 1, 7, and 8) were treated in essentially the same way as the three groups used by Williamsen et al. Specifically, both experiments included a hypnotic amnesia group, a waking simulation group, and a waking control group that received all instructions and suggestions orally and were assessed orally on Experimental and "Post"-Experimental Recall and on Recognition of Partial and Whole Words. To ascertain whether we had cross-validated the findings of Williamsen et al. we first compared the results we had obtained with our Groups 1, 7, and 8 with the results they had obtained (see Table 2).

Experimental and "Post-Experimental Recall (Table 2a). In both the Williamsen et al. experiment and our experiment, the scores on the recall tests were treated by analyses of variance and multiple comparisons were made among the three treatment means. These analyses yielded the following results:

1. With respect to Experimental Recall, in both experiments the simulators recalled none

TABLE 3

MEAN SCORES ON TESTS FOR AMNESIA

	1 Experimental recall	2 Recognition of partial words		3 Recognition of whole words	4 "Post"-Experimental recall	5 "Post"-Experimental subjective reports
		Critical words	Neutral words			
Spoken presentation (Groups 1–8)	3.9	3.8	1.2	4.2	5.1	1.8
Recorded-Written presentation (Groups 9–16)	3.4	3.6	1.5	4.2	5.4	1.2
Hypnotic induction (Groups 1–4 and 9–12)	4.0	3.9	1.4	4.2	5.3	1.1
No hypnotic induction (Groups 5–8 and 13–16)	3.2	3.5	1.3	4.2	5.2	1.8
Authoritative suggestions of amnesia, with hypnotic induction (Groups 1 and 9)	4.4	3.7	1.2	4.3	5.2	1.0
Authoritative suggestions of amnesia, without hypnotic induction (Groups 5 and 13)	3.5	4.3	1.6	4.7	5.4	2.0
Authoritative suggestions of amnesia, with and without hypnotic induction (Groups 1, 5, 9, and 13)	4.0	4.0	1.4	4.5	5.3	1.5
Permissive suggestions of amnesia, with hypnotic induction (Groups 2 and 10)	5.0	4.3	1.6	5.4	5.2	0.5
Permissive suggestions of amnesia, without hypnotic induction (Groups 6 and 14)	2.6	3.4	1.4	4.0	5.0	2.9
Permissive suggestion of amnesia, with and without hypnotic induction (Groups 2, 6, 10, and 14)	3.8	3.9	1.5	4.7	5.1	1.7
Instructions to simulate amnesia, with hypnotic induction (Groups 3 and 11)	1.3	3.3	1.5	1.8	5.2	2.4
Instructions to simulate amnesia, without hypnotic induction (Groups 7 and 15)	1.1	2.3	1.0	2.9	5.0	2.3
Instructions to simulate amnesia, with and without hypnotic induction (Groups 3, 7, 11, and 15)	1.2	2.8	1.2	2.4	5.1	2.4
No suggestions of amnesia, with hypnotic induction (Groups 4 and 12)	5.4	4.2	1.3	5.4	5.4	0.6
No suggestions of amnesia, without hypnotic induction (Groups 8 and 16)	5.8	4.2	1.1	5.1	5.5	0.0
No suggestions of amnesia, with and without hypnotic induction (Groups 4, 8, 12, and 16)	5.6	4.2	1.2	5.2	5.4	0.3

TABLE 4

2 × 2 × 4 ANALYSES OF VARIANCE OF SCORES ON TESTS FOR AMNESIA

Source	df	1 Experimental recall		2 Recognition of partial words				3 Recognition of whole words		4 "Post"-Experimental Recall		5 "Post"-Experimental subjective reports	
				Critical words		Neutral words							
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Spoken versus recorded-written presentation (A)	1	9.50	2.36	1.00		4.00	2.20	.17		1.18	1.90	12.84	2.30
Presence versus absence of hypnotic induction (B)	1	22.56	5.61*	4.69	1.66	.69		.17		.01		16.67	2.99
Types of suggestions (C)	3	123.00	30.60**	13.31	4.72*	.50		55.91	14.26**	1.15	1.85	25.99	4.67**
A × B	1	3.68		.25		3.36	1.85	.34		.56		.01	
A × C	3	6.42	1.60	1.16		2.05	1.13	9.02	2.30	.78	1.26	18.36	3.30*
B × C	3	12.88	3.20*	5.04	1.79	1.08		10.10	2.58	.32		16.52	2.96*
A × B × C	3	5.95	1.48	2.30		.56		7.49	1.91	.54		3.89	
Error	128	4.02		2.82		1.82		3.92		.62		5.57	

* $p < .05$.
** $p < .01$.

of the six critical words, the controls recalled practically all of the words, and the hypnotic group recalled significantly more words than the simulators and significantly fewer words than the controls.[4] Further, in both experiments the high susceptible hypnotic subjects recalled significantly fewer words than the low susceptible.[5]

2. With respect to "Post"-Experimental Recall, there were no significant differences in either our experiment or in the Williamsen et al. experiment between the hypnotic, simulating, and control groups, or between the high and low susceptible hypnotic subjects. In both experiments all groups "post"-experimentally recalled almost all (about five) of the six critical words.

Recognition of Partial Words (*Table 2b*). In both experiments, the simulators recognized significantly fewer of the *critical* words (when they were presented as partial words) than the hypnotic group and the controls. The latter two groups did not differ significantly from each other.

In both experiments, the three experimental groups did not differ significantly in recognition of the *neutral* (control) partial words.

In both experiments, the hypnotic and control groups (but not the simulators) recognized significantly more critical than neutral partial words.[6]

Recognition of Whole Words (*Table 2c*). In both experiments, the simulators recognized very few or none of the critical words (presented as whole words), the controls and the low susceptible hypnotic subjects recognized practically all of the words, and the high susceptible hypnotic subjects recognized significantly more words than the simulators and significantly fewer words than the controls.

---

[4] The .05 level of confidence is used as the criterion for significance in this and all other statistical analyses discussed in the present report.

[5] In the Williamsen et al. experiment, high and low susceptible refer, respectively, to subjects scoring 8 or higher and 4 or below on the 12-point Stanford Hypnotic Susceptibility Scale (Weitzenhoffer & Hilgard, 1959). In our experiment, high and low susceptible refer, respectively, to subjects passing 5 or more and less than 5 test-suggestions on the abbreviated 6-point Barber Suggestibility Scale.

[6] Latency of response, that is, the number of seconds required by the subject to recognize each word, was also measured. In both our experiment and in the Williamsen et al. experiment, the hypnotic and control groups recognized the critical words more quickly than the neutral words. The simulators did not show this reduction in response latency with respect to the critical words.

The results obtained in our experiment with Groups 1, 7, and 8 clearly cross-validate the results of the Williamsen et al. experiment. Although our results are practically identical to those of Williamsen et al., we cannot accept their implied conclusion that the presence or absence of "hypnotic trance induction" was an important variable in producing these effects. The results obtained with the additional 13 groups participating in our experiment strongly suggest a quite different interpretation, as indicated by the presentation that follows.

Overall Analyses of Results (2 × 2 × 4 Factorial)

The mean scores on each of the five dependent variables—Experimental Recall, Recognition of (Critical and Neutral) Partial Words, Recognition of Whole Words, "Post"-Experimental Recall, and "Post"-Experimental Subjective Reports—are presented in Table 3. These scores were treated by 2 × 2 × 4 analyses of variance and Duncan (1955) Range tests, which were used to make multiple comparisons among the means. The variance analyses, presented in Table 4, showed the following results with respect to each dependent variable.

Experimental Recall (Table 4, Column 1). The main effects of Presence versus Absence of Hypnotic Induction and Types of Suggestions were significant and the interaction between these variables was also significant. Further specification of these significant effects showed the following (see Table 3, also):

1. The significant effect of Presence versus Absence of Hypnotic Induction, interpreted in the light of the interaction between this variable and Types of Suggestions, indicated that the subjects who received the hypnotic induction procedure (hypnotic groups) and those who did not (waking groups) did not differ significantly (on Experimental Recall) when they received authoritative suggestions of amnesia, instructions to simulate amnesia, and no suggestions of amnesia. However, when permissive suggestions of amnesia were administered, the hypnotic groups showed *less* amnesia (verbalized more of the critical words) than the waking groups.

2. Overall, the simulating groups verbalized significantly fewer critical words than the groups given authoritative, permissive, or no suggestions of amnesia (controls). The groups given authoritative and permissive suggestions did not differ significantly from each other, and both of these groups verbalized significantly fewer words than the controls.

Recognition of (Critical and Neutral) Partial Words and Recognition of Whole Words (Table 4, Columns 2 and 3). The main effect of Types of Suggestions was significant with respect to recognition of the *critical* words both as partial words and as whole words. (The other main effects and interactions were nonsignificant.) Further delineation of the significant effect showed that the simulators recognized significantly fewer of the critical words than the groups given authoritative, permissive, or no suggestions of amnesia. The latter three groups did not differ significantly from each other.

With respect to recognition of *neutral* partial words, there were no significant main effects or interactions. All groups, irrespective of experimental treatments, recognized one to one and a half of the neutral words (see Table 3, Column 2).

All groups (simulators, controls, and groups given permissive and authoritative suggestions) recognized significantly more of the critical words than of the neutral words.[7]

"Post"-Experimental Recall. The variance analysis showed no significant main effects or interactions on "Post"-Experimental Recall (Table 4, Column 4). When told that the experiment was "over," all groups verbalized five to five and a half of the six words (see Table 3, Column 4).

"Post"-Experimental Subjective Reports—Subjective Scores (Table 4, Column 5).

---

[7] In the preceding section, in which we compared our results with the results of Williamsen et al., we stated that the simulating group which entered into the comparison ((Group 7) did not recognize significantly more critical words than neutral words. In the present analysis all four groups receiving one type of suggestion are combined. The combined simulating groups (Groups 3, 7, 11, and 15) recognized significantly more critical than neutral words.

There was a significant main effect for Types of Suggestions and also significant interactions of this variable with Spoken versus Recorded-Written Presentation and with Presence versus Absence of Hypnotic Induction.

The significant main effect of Types of Suggestions indicated that subjects in the control groups (no suggestions of amnesia) testified "post"-experimentally that they had forgotten significantly fewer words than subjects in the groups given authoritative, permissive, or simulation suggestions. The latter three groups did not differ significantly from each other.

The significant interaction of Types of Suggestions with Spoken versus Recorded-Written Presentation reflected the following. The simulators and the subjects given authoritative suggestions obtained significantly higher Subjective scores under the Spoken rather than under the Recorded-Written presentation. However, the Spoken and Recorded-Written presentations did not produce differences in subjective scores when permissive suggestions and no suggestions were administered.

The nature of the significant interaction of Types of Suggestions with Presence versus Absence of Hypnotic Induction can be discerned from Table 5 which presents the results of the Duncan Range tests. This table shows that hypnotic and waking subjects did not differ significantly from each other in subjective scores when given authoritative, simulation, or no suggestions of amnesia. However, when permissive suggestions ("Try to forget") were administered to hypnotic and waking subjects, the waking subjects rather than the hypnotic subjects obtained significantly higher subjective scores; that is, they claimed "post"-experimentally that they had experienced more amnesia.

Table 5 also shows the following: (*a*) the smallest subjective scores were obtained by the hypnotic groups given permissive and authoritative suggestions and by the hypnotic and waking control groups; (*b*) the highest subjective scores were obtained by the waking groups given permissive and authoritative suggestions and by the waking and hypnotic simulators.

The mean scores presented in Table 5 do not clarify an important question: of those subjects who ostensibly "forgot" more than one or two words during the experiment, what proportion testified "post"-experimentally that they had actually forgotten them? To answer this question, Table 6 presents the number and percentage of subjects in the various experimental groups who (*a*) "forgot" three or more of the six critical words on the Experimental Recall test, (*b*) verbalized the "forgotten" words on the "Post"-Experimental Recall test, and then (*c*) testified ("Post"-Experimental Subjective Reports) that they had forgotten the words either "completely" or "partially." The upper part of the table shows that very few subjects (0% to 11%) in any of the experimental groups claimed "complete" amnesia for three or more words. The bottom part of the table shows that, with the exception of one group, the majority of subjects in each of the experimental groupings did *not* claim either "complete" or "partial" amnesia. The exception here was the waking groups given permissive suggestions of whom half claimed either "complete" or "partial" amnesia.

"Good"-"Poor" Hypnotic Subjects and

TABLE 5

RESULTS OF DUNCAN RANGE TESTS APPLIED TO MEAN SUBJECTIVE SCORES

	Groups							
	Authoritative suggestions		Permissive suggestions		Instructions to simulate		No suggestions	
	1 & 9	5 & 13	2 & 10	6 & 14	3 & 11	7 & 15	4 & 12	8 & 16
Hypnotic induction	1.0_{bcd}	—	0.5_{cd}	—	2.4_{ab}	—	0.6_{cd}	—
No hypnotic induction	—	2.0_{abc}	—	2.9_{a}	—	2.3_{ab}	—	0.0_{cd}

Note.—Means containing a common subscript do not differ from each other at the .05 level of confidence by Duncan Range test.

TABLE 6

Number (and Percentage) of Subjects Testifying "Post"-Experimentally that They Had Forgotten Three or More Words Either "Completely" or "Partially"

	Authoritative suggestions	Permissive suggestions	Instructions to simulate	No suggestions
"Complete" amnesia				
Hypnotic induction	0 (0%)	0 (0%)	1 (5.5%)	1 (5.5%)
No hypnotic induction	1 (5.5%)	1 (5.5%)	2 (11.1%)	0 (0%)
Either "complete" or "partial" amnesia				
Hypnotic induction	3 (16.7%)	1 (5.5%)	8 (44.4%)	1 (5.5%)
No hypnotic induction	5 (27.7%)	9 (50.0%)	6 (33.3%)	0 (0%)

Note.—$N = 18$ in each cell.

Response to Suggestions of Amnesia. Of the 72 subjects tested under the hypnotic induction condition, 24 (33%) were rated as "excellent" hypnotic subjects (passing all six of the test suggestions on the abbreviated BSS), 8 (11%) were rated as "good" hypnotic subjects (passing five of the six test suggestions), 21 (29%) were rated as "fair" (passing three or four), and 19 (26%) were rated as "poor" (passing two or less). To ascertain whether the "better" hypnotic subjects were the most responsive to suggestions of amnesia, Pearsonian correlations were computed between (*a*) scores on the BSS and scores on the Experimental Recall test and (*b*) scores on the BSS and subjective scores (scores on the "Post"-Experimental Subjective Reports test).

The correlations are presented in Table 7. Inspection of the table shows that, with respect to hypnotic subjects given authoritative suggestions of amnesia, the higher their scores on the BSS the more they manifested amnesia, that is, the fewer the number of words they verbalized on the Experimental Recall test and the more they claimed "post"-experimentally to have experienced "complete" or "partial" amnesia. The correlations with respect to the hypnotic subjects given permissive suggestions of amnesia are in the same direction but fail to reach significance ($p < .10$, one-tailed test). The correlations for the simulators and the controls are far from significant.

DISCUSSION

Of the three independent variables in this experiment, two of these, Spoken versus Recorded-Written Presentation and Presence versus Absence of Hypnotic Induction, did not yield very powerful effects, whereas the third, Types of Suggestions, was a highly potent variable. The effects of each of these independent variables will be discussed in turn.

Spoken versus Recorded-Written Presentation

Overall, there were no significant differences in response to the tests for amnesia between (*a*) the groups receiving all suggestions, instructions, and tests for amnesia orally and (*b*) the groups receiving the same suggestions and instructions by means of a tape recording and the tests for amnesia by means of a written questionnaire. This overall nonsignificant main effect is in line with a previous study (Barber & Calverley, 1964a) which similarly failed to find differences in response to test suggestions when the suggestions were given orally and by means of a recording.

The overall finding, however, requires a qualification with respect to one of the five dependent variables ("Post"-Experimental Subjective Reports). The Spoken presentation was more effective than the Recorded-Written presentation in producing high subjective scores in the simulators and in the groups given authoritative suggestions but not in the controls or in the groups given permissive suggestions. Further careful studies are needed to cross-validate this interaction and to clarify its significance.

Presence versus Absence of Hypnotic Induction

With one exception, subjects receiving a hypnotic induction (hypnotic groups) and those not receiving an induction (waking groups) did not differ in amnesic performance, as follows:

1. Hypnotic controls and waking controls did not differ significantly in response to any of the tests for amnesia. This outcome is

consistent with a series of recent experiments (Barber, 1958, 1962b, 1965) which indicate that very few if any hypnotic subjects forget the experimental events when amnesia is not directly or tacitly suggested.

2. Hypnotic simulators did not differ from waking simulators in response to any of the tests for amnesia. This outcome cannot be compared with that of other studies; in no previous study known to us were instructions to simulate amnesia given to both hypnotic and waking groups.

3. Hypnotic and waking subjects who received authoritative suggestions ("You will not remember!") did not differ significantly in response to any of the tests for amnesia. This outcome is in line with a series of previous experiments (Barber & Calverley, 1962, 1963b, 1963c; Barber & Glass, 1962) which failed to find significant differences between hypnotic and waking subjects in response to peremptory suggestions that they would forget one of the experimental events.

4. When permissive suggestions ("Try to forget") were administered, the hypnotic groups showed *less* amnesia than the waking groups (verbalizing more critical words on Experimental recall and obtaining smaller subjective scores). This surprising outcome requires cross-validation. If this finding is confirmed in replicative studies, it may be possible to relate it to previous investigations which indicated that *some* types of permissive suggestions—for example, permissive suggestions of deafness ("Try not to hear"), permissive suggestions for enhanced endurance ("Try your best to hold the weight for a much longer time than you did before"), permissive suggestions of analgesia ("Try to imagine that the [pain-producing] stimulus is pleasant"; Barber & Calverley, 1964c, 1964d; Barber & Hahn, 1962)—are remarkably effective with waking subjects.

Types of Suggestions

The most potent independent variable in the present experiment was Types of Suggestions. Hypnotic and waking groups given no suggestions (controls) did not show amnesia. In contrast, hypnotic and waking groups given authoritative or simulation suggestions and waking groups given permissive suggestions showed some apparent amnesia on Experimental Recall and "Post"-Experimental Subjective Reports. In addition, the hypnotic and waking simulators also showed some apparent amnesia on Recognition of Partial and Whole Words.

The most surprising outcome of the present experiment was that, in general, apparent amnesia could be produced most effectively by administering simulation instructions to waking and hypnotic subjects and permissive suggestions to waking subjects. These groups tended to manifest more amnesia than the hypnotic groups given permissive suggestions and the hypnotic and waking groups given authoritative suggestions (see Table 3, Columns 1 and 5).

We had assumed that subjects asked to simulate amnesia would state post-experimentally that they did not forget. The present data sharply contradict this assumption. A substantial proportion (33% to 44%) of hypnotic and waking subjects who had been asked to simulate amnesia testified post-experimentally, when they were no longer simulating, that they had "completely" or "partially" forgotten at least three of the six critical words (see Table 6). This outcome does not contradict any previous studies known to us. We found no prior studies in which subjects were instructed to simulate amnesia and then were asked postexperimentally if they had actually forgotten. Furthermore, in those studies (e.g., Shor, 1959) in which subjects were asked to simulate other hypnotic phenomena—for example, hypnotic analgesia—those simulators who stated that they had actually experienced that which they were supposed to simulate were deliberately excluded from the analysis of the data and were not mentioned in the discussion of results.

Further Considerations

Several additional findings require emphasis:

The correlations presented in Table 7 indicate that those hypnotic subjects who are the most responsive to test suggestions of arm levitation, body immobility, and the other test suggestions included in the BSS also tend to be the most responsive to peremptory or permissive suggestions of amnesia. However, hypnotic subjects who have been asked to simulate amnesia perform their role satisfactorily (i.e., manifest amnesic behavior)

TABLE 7

CORRELATIONS FOR HYPNOTIC SUBJECTS BETWEEN SCORES ON ABBREVIATED BARBER SUGGESTIBILITY SCALE (BSS) AND EXPERIMENTAL RECALL AND BETWEEN BSS AND SUBJECTIVE SCORES

	BSS-experimental recall	BSS-subjective scores
Authoritative suggestions of amnesia (Groups 1 and 9)	−.60*	+.48*
Permissive suggestions of amnesia (Groups 2 and 10)	−.34	+.36
Instructions to simulate amnesia (Groups 3 and 11)	+.21	+.20
No suggestions of amnesia (Groups 4 and 12)	−.02	+.09

* $p < .05$.

regardless of whether or not they are responsive to suggestions of arm levitation, body immobility, and so on.

If a recorded-written presentation had not been used in the present experiment, one could argue that the results might be due to bias on the part of the experimenter. That is, it might have been argued that the experimenter consciously or unconsciously desired and expected more amnesic responses from some groups than from others and that his desires and expectations, by affecting the manner and the tone in which he administered the suggestions and assessed the amnesia, influenced the experimental results (Barber & Calverley, 1964b). However, by using a recorded-written presentation with half of the subjects, this argument becomes invalid. Subjects given the suggestions by means of a tape recording and the tests for amnesia by means of a written questionnaire gave essentially the same results as those given the suggestions and tests for amnesia orally.

Williamsen et al. (1965) implied that the presence or absence of "hypnotic trance induction" was an important variable in producing the differences in performance among hypnotic, simulating, and control groups. Although our findings with Groups 1, 7, and 8 unequivocally confirmed their results, the findings obtained with the additional groups participating in our experiment controvert their implied conclusion, as follows: (*a*) *In most instances, "hypnotic trance induction" does not significantly affect performance.* Hypnotic and waking subjects told peremptorily, "You will not remember," did not differ on the tests for amnesia. Hypnotic and waking subjects instructed to simulate amnesia also did not differ in performance. Furthermore, hypnotic and waking controls performed similarly on the tests. (*b*) *When "hypnotic trance induction" affects performance, it may produce less amnesia than would be found without it.* When hypnotic and waking subjects were told, "Try to forget," the hypnotic subjects showed *less* amnesia than the waking subjects. (*c*) *The important factors in producing differences in performance on tests for amnesia are whether or not suggestions of amnesia are administered and the types of suggestions administered.* Irrespective of the presence or absence of "hypnotic trance induction," subjects differed on the tests for amnesia depending on whether they were given suggestions to simulate, authoritative or permissive, or no suggestions of amnesia.[8]

Denotations of the Term "Amnesia"

The global construct "amnesia" has been used to conceptualize behaviors exhibited in the present experiment. To understand the import of the findings it is necessary to specify what this construct denotes. The data summarized below suggest that the "amnesia" exhibited in the experiment may have consisted primarily of unwillingness to verbalize the critical words rather than "actual forgetting" of the words.

When told "the experiment is over," the "amnesic" subjects in *all* groups readily verbalized, that is, easily remembered, the critical words.

---

[8] Two additional points can be noted parenthetically: (*a*) It appears that amnesia can be produced as easily by suggestions given to subjects who are compelled to participate in the experiment as in those who volunteer. We used "captive" females of ages 17–25 whereas Williamsen et al. worked with females of similar ages who volunteered for hypnosis. As pointed out above, the three experimental groups in the Williamsen et al. study and our three similarly treated groups (Groups 1, 7, and 8) responded in essentially the same way to the tests for amnesia. (*b*) It also appears that simulators will fulfill their role-assignment satisfactorily regardless of whether the same experimenter (as in the present study) or different experimenters (as in the Williamsen et al. study) are used to instruct them and to assess their performance (cf., Orne, 1959).

The "amnesic" subjects in the hypnotic and waking groups that had received authoritative or permissive suggestions showed little if any evidence of amnesia when they were tested indirectly by means of the recognition tests.

The "amnesic" subjects in all groups *recognized more critical words than neutral (control) words* when they were presented as partial words. This outcome appears to indicate that the ostensibly "amnesic" subjects in *all* groups were quite "aware" of the critical words and had not "actually forgotten" them.

The overwhelming majority of "amnesic" subjects testified "post"-experimentally that they "forgot" only "partially" in that they "thought of the critical words at certain times, off and on, during the experiment" (see Table 6).

Very few "amnesic" subjects testified "post"-experimentally that they "forgot completely" in that they "did not think about the critical words at all during the experimen and the words did not enter their mind" (see Table 6). It may be that these few subjects did not "think about the critical words at all." On the other hand, these subjects may have stated that they "did not think about the words" in order to please the experimenter or to try to do what they thought was expected of them. The latter possibility appears more likely in that, during the experiment, the subjects had verbalized ("thought about") most of the critical words when they were presented as partial words or as whole words and, more important, they recognized and verbalized more of the *critical* words than of the neutral (control) words.

REFERENCES

BARBER, T. X. Hypnosis as perceptual-cognitive restructuring: II. "Post-hypnotic" behavior. *Journal of Clinical and Experimental Hypnosis,* 1958, **6,** 10–20.

BARBER, T. X. Experimental controls and the phenomena of "hypnosis": A critique of hypnotic research methodology. *Journal of Nervous and Mental Disease,* 1962, **134,** 493–505. (a)

BARBER, T. X. Toward a theory of hypnosis: Posthypnotic behavior. *Archives of General Psychiatry,* 1962, **7,** 321–342. (b)

BARBER, T. X. Experimental analyses of "hypnotic" behavior: A review of recent empirical findings. *Journal of Abnormal Psychology,* 1965, **70,** 132–154.

BARRBER, T. X., & CALVERLEY, D. S. "Hypnotic behavior" as a function of task motivation. *Journal of Psychology,* 1962, **54,** 363–389.

BARBER, T. X., & CALVERLEY, D. S. "Hypnotic-like" suggestibility in children and adults. *Journal of Abnormal and Social Psychology,* 1963, **66,** 589–597. (a)

BARBER, T. X., & CALVERLEY, D. S. The relative effectiveness of task motivating instructions and trance induction procedure in the production of "hypnotic-like" behaviors. *Journal of Nervous and Mental Disease,* 1963, **137,** 107–116. (b)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of task motivating instructions and attitudes toward hypnosis. *Journal of Abnormal and Social Psychology,* 1963, **67,** 557–565. (c)

BARBER, T. X., & CALVERLEY, D. S. Comparative effects on "hypnotic-like" suggestibility of recorded and spoken suggestions. *Journal of Consulting Psychology,* 1964, **28,** 384. (a)

BARBER, T. X., & CALVERLEY, D. S. Effect of *E*'s tone of voice on "hypnotic-like" suggestibility. *Psychological Reports,* 1964, **15,** 139–144. (b)

BARBER, T. X., & CALVERLEY, D. S. Experimental studies in "hypnotic" behaviour: Suggested deafness evaluated by delayed auditory feedback. *British Journal of Psychology,* 1964, **55,** 439–446. (c)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of "hypnotic" behaviour: Enhancement of strength and endurance. *Canadian Journal of Psychology,* 1964, **18,** 156–167. (d)

BARBER, T. X., & GLASS, L. B. Significant factors in hypnotic behavior. *Journal of Abnormal and Social Psychology,* 1962, **64,** 222–228.

BARBER, T. X., & HAHN, K. W., JR. Physiological and subjective responses to pain producing stimulation under hypnotically-suggested and waking-imagined "analgesia." *Journal of Abnormal and Social Psychology,* 1962, **65,** 411–418.

BARBER, T. X., KARACAN, I., & CALVERLEY, D. S. "Hypnotizability" and suggestibility in chronic schizophrenics. *Archives of General Psychiatry,* 1964, **11,** 439–451.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics,* 1955, **11,** 1–42.

FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *Journal of Abnormal and Social Psychology,* 1938, **33,** 453–475.

MARCUSE, F. L. *Hypnosis: Fact and fiction.* Baltimore: Penguin Books, 1959.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology,* 1959, **58,** 277–299.

SHOR, R. Explorations in hypnosis: A theoretical and experimental study. Unpublished doctoral dissertation, Brandeis University, 1959.

TROFFER, S. A., & TART, C. T. Experimenter bias in hypnotist performance. *Science,* 1964, **145,** 1330–1331.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

WILLIAMSEN, J. A., JOHNSON, H. J., & ERIKSEN, C. W. Some characteristics of posthypnotic amnesia. *Journal of Abnormal Psychology,* 1965, **70,** 123–131.

(Received March 9, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 108-114

VISUAL VERSUS TACTUAL RECEPTOR PREFERENCE IN NORMAL AND SCHIZOPHRENIC CHILDREN[1]

ERIC SCHOPLER[2]

University of Chicago

Receptor preferences between visual or tactual stimuli were measured in schizophrenic and normal children. 4 pairs of standardized choice situations between visual or tactual stimuli were used, with the time of engagement with either stimulus constituting the preference measure. Schizophrenic children between the ages of 7 and 9 were compared with same aged normal children. A group of retardates were used for MA control. Schizophrenics were significantly lower in visual preference than the same aged normals, and an age trend for increased visual preference was found in the normal sample. Retardates showed greater visual preference than schizophrenics of comparable MA. Some implications for theory of schizophrenia were discussed.

The problem of hierarchies in receptor usage was identified by Sherrington (1906) who postulated a transition from near receptor dominance in lower organisms to distance receptor dominance in the higher forms. A parallel sort of ontogenetic development has been suggested by Schachtel (1959), who described the infant's change from the autocentric mode of perception, utilizing primarily the near receptors such as touch, to the allocentric mode, involving such distance receptors as vision. This shift is consistent with Piaget's (1952) developmental sequences of the sensory-motor period during which change from reflex behavior to increased goal directedness occurs. This change corresponds with Werner's (1957) discussion of development from fusion to differentiation. On the empirical side, a few studies suggest development from touch to vision. Renshaw et al. (1930) found younger children to be more accurate in tactual-kinesthetic localization than they were in visual localization, while the pattern was reversed in older children. Zaporozhets (1961) reported that exploration of novel situations by young children changed from tactual to visual modalities with age. Tarakanov and Zinchenke (1960) reported better recall after tactual than after visual exploration in young children. The above literature suggests that, in normal development, a progressive shift from near receptor preference (touch) to distance receptor preference (vision) occurs with increasing maturity.

Various theories of childhood schizophrenia (Ekstein, Bryant, & Friedman, 1958) emphasize that such children display regression or lag in certain developmental functions. Clinical observations from several sources (Goldfarb, 1956, 1961; Mahler, 1952; Schopler, 1965) suggest that schizophrenic children make poor use of their visual and auditory acuity, while they make excessive use of touch, taste, and smell for orienting to their environment. In fact, some recent therapeutic efforts have been oriented primarily to tactual contact with such children (Des Lauriers, 1962; Schopler, 1962; Waal, 1955). In view of these observations, it may be hypothesized that schizophrenic children will show the type of receptor preferences typically associated

---

[1] This study is based on a doctoral dissertation submitted to the University of Chicago, and was supported in part by United States Public Health Service Fellowship MPM 14, 386-R2. The author is grateful to Sheldon H. White for his encouragement and advice during all phases of this research. Comments of Lawrence Kohlberg and Robert Van de Castle improved the published version.

Subjects and testing facilities were made available through the generous cooperation of the Peoria State Hospital, Walter J. Garre, Superintendent; The Research Center for the Study of Childhood Schizophrenia, Indiana University, M. K. De Myer, Director; Treatment and Research Center for Childhood Schizophrenia, I. Yahalom, Director; the Social Psychology Laboratory, University of Chicago, R. L. Strodtbeck, Director; University of Chicago Laboratory School, Robert E. Newman, Principal of Primary Grades; and the Chicago Day School, M. Andrews, Principal.

[2] Now at the University of North Carolina, School of Medicine, Department of Psychiatry, Chapel Hill, North Carolina.

TABLE 1

CHARACTERISTICS OF SCHIZOPHRENIC AND RETARDED SUBJECTS

Subject	Age range (yrs. & mos.)	Sex		Mean IQ	Mean MA (yrs. & mos.)	Facility	
		Boys	Girls			In-patient	Out-patient
Schizophrenics ($N = 30$)	5–6 to 6–5	3	—	54	3–8	2	1
	6–6 to 7–5	4	2	92	6–4	2	4
	7–6 to 8–5	7	4	84	6–6	8	3
	8–6 to 9–5	8	2	87	8–2	5	5
M	7–10	—	—	84	6–8	—	—
Retardates $N = (15)$							
M	10–2	—	—	66	6–7	—	—

with normal children of a younger age level. If the hypothesized differences between normal and schizophrenic children occur, they may be a function of schizophrenic factors or a function of the schizophrenics' lower MA.

The present study was carried out to investigate whether empirical support could be found for the hypothesized receptor preferences of normal and schizophrenic children when the effects of MA were controlled.

METHOD

Subjects

Schizophrenic sample. The 30 children in this sample, see Table 1, had in common the diagnostic label of schizophrenic or psychotic, as recorded in their clinic charts, according to psychiatric or psychological judgment. IQ scores ranged from 50 to 129 and were obtained from either the WISC or Stanford-Binet Form L. None of these children had any gross defects in visual acuity according to observations of staff members responsible for their care.

Normal sample. The normal group consisted of 90 children, 47 boys and 43 girls, drawn from nursery groups and the first four grades of the University of Chicago Lab school. Ten children were drawn randomly from each of the seven grade divisions (two nursery groups, kindergarten, and the first four grades). An additional 20 *S*s were selected from the second and third grade because the greatest number of schizophrenics available for comparison were in this age range. Individual IQ scores were not available for children in the first four grades. Based on subsequent grades, mean IQs for this group were in the superior range, and the children came primarily from professional families.

Retarded sample. By comparing a group of mentally retarded children of MA similar to that of the schizophrenics, the effects of both MA and schizophrenic factors could be evaluated. From a school for educable retarded boys, a group of 15 *S*s were selected, diagnosed as nonspecific, familial types of retardates by the school's testing staff. As shown in Table 1, the children in this group had MAs comparable to those of the schizophrenic group. MAs were computed from IQ scores on the Binet Test Form L M.

Procedures

Four standardized situations were developed to permit a comparison between the child's interest in stimuli yielding to tactual exploration and his interest in stimuli yielding to visual exploration. Because schizophrenic children tend to follow verbal instructions poorly, if at all, the tests were designed for minimal reliance on verbal instructions. The four preference situations are described below.

The Slides Test. In this test the visual instrument consisted of a Kodak Carousel slide projector, encased in a masonite box, set to change slides automatically at 20-second intervals. In order for the color slides to be projected on a 2 × 3 foot screen, a switch had to be pressed which in turn opened the necessary shutter. The slides were colored pictures either of animals or scenery; they excluded any human figure so that avoidance of human beings would not be confounded with visual avoidance. The tactual component for this situation consisted of a vibrator enclosed in a grey 12 × 5 × 4 inch plywood box. When a lever on the box was pressed a moderate hand vibration was produced. The visual and tactual components were so situated in the testing room that they were not accessible simultaneously. The child was introduced to each component in alternated order with the directions "Here are two things for you to play with. When you push this button you see a picture on the screen. Try it!" (One picture change was shown.) "If you push this button you will feel it in your hand. Try it!" If the child showed reluctance to push either button, *E* demonstrated first with his hand and then, if necessary, with the child's hand. The child was then taken to a point equidistant from both switches and was told that he could try both until the game was over and was asked which he wanted to do first. In the event a child remained with his first choice for 5 minutes, he was directed to the alternate possibility with the qualification that he could continue with either one thereafter, for as long as he wished. The maximum length of time allotted to both components of this test was 12 minutes and was used only by a few children. Both slide and vibrator were connected to time clocks which recorded the length of time each switch was closed.

The Kaleidoscope Test. For this procedure the visual-tactual choice was between a kaleidoscope and a piece of dark-hued putty, selected to resist the forming of different shapes but to yield primarily to exploration of its texture. The kaleidoscope was demonstrated when unfamiliar to the child. The order of presenting either kaleidoscope or putty first was alternated between *S*s. As with the previous procedure, if the child stayed with the first stimulus for 5 minutes, he was interrupted and given the other one. He was timed on his exploration of each component and could terminate his exploration whenever he wished, up to a maximum of 10 minutes for this test.

The Blocks Test. The child was seated at a table on which 20 blocks had been randomly placed, all within his reach. These blocks consisted of 2 sets of 2½-inch wooden cubes, 10 in each set. One set of 10 blocks was covered with different color designs and pictures but was of uniform texture. The second set, on the other hand, was covered with different textures, such as fur and sandpaper, but with minimal variation in color. The child was told that he might inspect these blocks in any way he wished until told that time was up. This exploratory situation offered observational measures of his interest in the various blocks. The rating of these observations were to be made according to the mode of exploration. Thus, when a child explored the texture blocks visually, by turning them, this was timed for visual interest. If, on the other hand, he explored a color block by feeling or rubbing, this was scored as tactual time. Coordinated activity, such as building, was scored equally for touch and vision. Three minutes were allotted to the exploration of these blocks.

The Disc Test. For the visual component, seven toy animals about 10 inches long were mounted on a disc of 30-inch diameter which was rotated by a motor at one RPM. The disc was enclosed in a 32 × 32 inch black box with two openings lighted by a 40-watt bulb, through which the rotating toy animals could be observed. Two black curtains were placed so that the figures could only be seen when the observer stood in front. The disc apparatus was placed against the wall on a table, where it could not be reached by a child from the floor. The tactual component of this test was a 72 × 14 × 14 inch enclosed, plywood box, divided into five compartments. Each compartment was accessible through an opening with cloth sleeves to prevent visual inspection. A toy animal similar to the ones on the disc was fastened to the bottom of each compartment. The child was introduced to this situation with the following explanation: "Here are two things. You can watch those toy animals go around, or you can feel the toys in the box with your hands." If he was reluctant, *E* demonstrated with his own hand. The order of introducing the visual or tactual components first was alternated between *S*s. After this introduction the child was taken to a point in the room equidistant from both components. He was reminded that he could explore either one until the game was over. The child's exploration of each component was timed. If he spent 5 minutes on one without moving away, *E* suggested that he try the other one and return whenever he wished. This was done in case the child did not understand or remember directions, or in case he perseverated on one stimulus. The *S* could then engage himself with either stimulus until he wished to stop. The maximum time of 10 minutes for this test was used by only a few children.

RESULTS

The data for this study were collected on four tasks and included two kinds of measures:[3] (*a*) the absolute length of time *S* remained engaged with either visual or tactual stimuli on each of the four procedures; (*b*) the proportion of time *S* remained engaged with the visual or tactual stimulus on each procedure. The proportion is expressed as percentage of visual time in the following analysis.

Age Trends for Normal Children

With the normal groups an analysis of variance was made between boys and girls for total time spent on all four procedures ($F = 2.31$, and 1/89 *df*, $p > .10$) and also the total percentage visual time on all four procedures ($F = 1.86$, 1/89 *df*, $p > .10$). Although girls tended to score higher than boys, this difference was not significant. Hence boys and girls were not separated for the age-trend analysis.

Figures 1 and 2 include the normal age trends on all four procedures.

Figure 1 shows that the older children spent more absolute visual time than did the younger children. There was, however, no age difference for tactual time spent. Figure 2 shows that the proportion of visual time spent is greater for the older children than for the younger. The significance of these age trends was evaluated using a computer programed regression analysis.[4] The correlation between the total percentage visual time and age over all four tests was .30 (89 *df*, $p < .005$). Similarly the correlations between age and absolute time spent on all four tests was .35 (89 *df*, $p < .0005$). There was no significant correlation between absolute tactual time and age. These trends, summarizing tactual and visual interest on all four procedures show that length of visual interest increased with age. Although length of tactual interest did not decrease with age, the proportion of tactual time decreased.

---

[3] The questions investigated in this study were primarily concerned with the proportion of time the subject chose to spend on each test with either modality. These mean percentages are summarized in Figure 2. However, if only the proportion measure were used, the changes shown in Figure 2 could be due to a decrease in absolute tactual time only, or an increase in absolute visual time. Although the two measures are not independent they are both included to show what trends in absolute time Figure 1 made for the changes in the proportion of visual time.

A similar regression analysis was performed for each of the four tests separately. On the Slide Test, the older normal children spent more time looking at slides than did the younger, ($r = .38$, 89 *df*, $p < .0001$). There was no significant age difference in the absolute time spent feeling the vibrator. The relative preference or percentage of visual time against tactual time increased with age ($r = .35$, 89 *df*, $p < .0005$). On the Kaleidoscope Test, as on the Slide Test, the older children spent more absolute time on the visual stimulus than did the younger ($r = .38$, 89 *df*, $p < .0001$). No significant age difference was found for absolute time spent on the putty. Percentage of visual time increased with age ($r = .35$, 89 *df*, $p < .0005$).

It was noteworthy that most children spent more time on the slides than on the vibrator, while conversely on the Kaleidoscope Test more absolute time was spent on the putty than on the kaleidoscope. Even without inherent equivalence of interest in the two sets of visual and tactual stimuli, the trend for increased visual preferences was clearly present in both tests. On the Blocks Test the age trend for visual exploration was ($r = .22$, 89 *df*, $p < .05$), and there was no significant trend for tactual time. There was no age trend on the Disc Test.

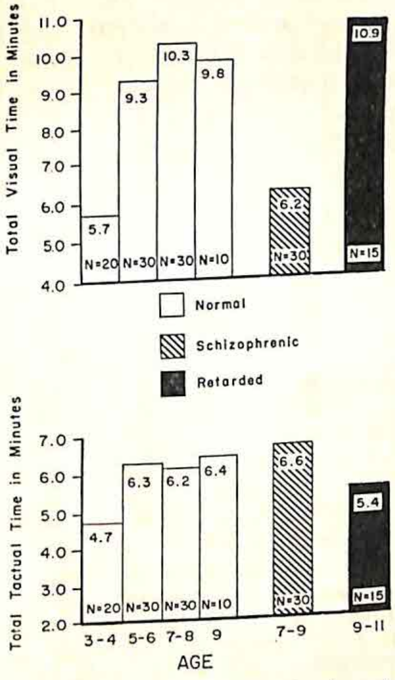


FIG. 1. Total time spent with visual and tactual stimuli in all four procedures.

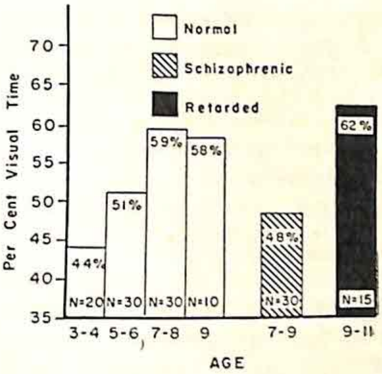


FIG. 2. Total percentage of visual versus tactual time spent on all four procedures.

Comparison between Normal and Schizophrenic Children

For the schizophrenic children no significant differences were found on total preference scores either between sexes or between in-patients and out-patients. Hence they were combined in the computer program. For those measures in which the regression analysis had revealed an age trend for the normals, the age factor was controlled through the computation of a *C*-score for each schizophrenic *S*.

---

[4] The computer program was derived from Sterling and Pollack (1963) and is described in Schopler (1964).

TABLE 2

MEAN DEVIATIONS OF SCHIZOPHRENICS FROM NORMALS WITH AGE FACTOR CONTROLLED

Variable	Schiz. *M* *C*-scores	*SE*	$p \leq .02$	Expected direction
Slide Test				
Visual time	.65	.21	Yes	Yes
Tactual time	−.22	.13	No	No
% Visual time	.87	.30	Yes	Yes
Kaleidoscope test				
Visual time	1.34	.08	Yes	Yes
Tactual time	.61	.13	Yes	No
% Visual time	.91	.15	Yes	Yes
Blocks Test				
Visual time	1.04	.18	Yes	Yes
% Visual time	.71	.34	Yes	Yes
Disc Test				
Visual time	.46	.18	Yes	Yes
Tactual time	.03	.24	No	No
% Visual time	.69	.26	Yes	Yes
All Four Tests				
Total Visual time	.86	.17	Yes	Yes
Total Tactual time	.32	.19	No	No
Total % Visual time	.91	.26	Yes	Yes

Note.—In the computer program when the mean standard score is positive it indicates the schizophrenics had a lower raw score than the normals. When it is negative, it indicates a higher raw score for the schizophrenics.

This *C*-score consisted of the best estimate of the mean scores obtained by a given normal age group, minus the observed score of each schizophrenic *S* in that age group, divided by the standard error of the normals' scores about their mean. Thus each schizophrenic *S* was compared with the same-aged normal children. The difference between schizophrenic and normal children is shown in Figures 1 and 2, and summarized in Table 2. This table shows that the *C*-scores of the schizophrenics were significantly lower than those of the same-aged normal group, on all visual measures. This confirmed the expectation that schizophrenic children will show less visual preference than same-aged normal children. The last column in Table 2 shows that the schizophrenic children did not spend more absolute time on tactual exploration than did the normals, underscoring the differences found between the two groups on all the visual measures.

Comparison of Schizophrenics with Retardates

The above differences between schizophrenics and normals were obtained with chronological age controlled. This left MA uncontrolled, and lower for the schizophrenics than the normals. To determine the effect of MA on preference for visual and tactual receptors, a group of retarded boys was compared with a group of younger boys, selected for their similarity in MA. Table 3 shows that the retardates spent more time on both visual and tactual measures for all four procedures than did the normal children of the same MA range. However, the differences were significant on three visual measures only—that is, on absolute time spent on kaleidoscope, on percentage of time spent with kaleidoscope, and on absolute time spent with the slides. Since the retardates showed as much or more visual preference than a normal group of younger children of similar MA, it would suggest that a lower IQ, also present in the schizophrenics, did not in itself account for the weaker visual preference of the schizophrenics.

The retardate boys were compared with schizophrenic boys of comparable MA (see Table 4). The retardates had a mean MA of 6–7, in a range from 5–5 to 8–5, while the schizophrenic boys had a mean MA of 6–9 in a range from 3–1 to 10–0. From the similarities between retardates and normals, it was to be expected that schizophrenics would have lower visual scores than retardates with similar MA. Table 4 shows that the schizophrenics had consistently lower visual scores than

TABLE 3

MEAN DIFFERENCES BETWEEN NORMAL AND RETARDED BOYS OF COMPARABLE MENTAL AGE

Variable	*M* for retardate ($N = 15$)	*M* for normal ($N = 15$)	*t* value
Slide Test			
Visual time	5.3	3.0	2.80*
Tactual time	1.9	1.0	1.61
% Visual time	78.0	72.0	1.03
Kaleidoscope Test			
Visual time	1.8	0.7	2.75*
Tactual time	1.8	1.6	.43
% Visual time	52.0	37.0	2.08*
Blocks Test			
Visual time	1.7	1.6	.82
% Visual time	54.0	54.0	.98
Disc Test			
Visual time	2.0	1.7	.62
Tactual time	1.1	.8	.53
% Visual time	67.0	63.0	1.48
All Four Tests			
Visual time	10.9	10.1	1.69
Tactual time	5.4	6.2	1.74
% Visual time	62.0	61.0	.70

* $p \leq .02$, 28 *df*.

TABLE 4

MEAN DIFFERENCES BETWEEN SCHIZOPHRENIC AND RETARDED BOYS OF COMPARABLE MENTAL AGE

Variable	M for retardate ($N = 15$)	M for schizophrenic ($N = 19$)	t value
Slide Test			
Visual time	5.3	3.1	2.56**
Tactual time	1.9	2.0	0.15
% Visual time	78.0	60.0	2.65**
Kaleidoscope Test			
Visual time	1.8	0.6	3.17***
Tactual time	1.8	1.4	1.02
% Visual time	52.0	32.0	3.52***
Blocks Test			
Visual time	1.7	1.3	1.94*
% Visual time	54.0	43.0	3.52***
Disc Test			
Visual time	2.0	.9	3.05***
Tactual time	1.1	1.0	.20
% Visual time	67.0	44.0	3.71***
All Four Tests			
Visual time	10.9	6.1	3.21***
Tactual time	5.4	46.4	1.64
% Visual time	62.0	48.0	3.55***

* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .005$.

retardates, but they were not different from each other on tactual scores.

The comparative effect of MA on the receptor preferences was also examined by comparing the correlation coefficients between MA and the receptor measures. These findings are reported in greater detail elsewhere (Schopler, 1964). The schizophrenics showed a positive r between visual measures and MA on 8 out of 10 measures, while for the retardates 5 out of these 10 showed a negative correlation. For the schizophrenics r was negative between MA and 3 out of 4 tactual measures while the retardates showed a positive correlation on 3 out of these 4. These differences, discussed below, may suggest a qualitatively different use of the same sense modality by the two groups.

DISCUSSION

The results supported the expectation that normal children will show a developmental tendency for increased visual preference with increased age. These age trends were significant although there was considerable variance among the mean preference scores, especially in the older end of the age range. This increasing variance of visual scores in the older children may be the beginning of a downward trend in visual preference as the child approaches adolescence. The most striking increase in visual interest occurred between the age of 3 to 6. It may be anticipated that on further study an even steeper portion of the visual interest growth curve may be associated with the first 3 years not covered by this study.

The expected decrease in tactual time with increased age was not supported by the findings. While there was an age trend for absolute visual time, there was none for absolute tactual time. This meant that the age trend for the increase in percentage of visual time was due to the increase in absolute visual time. These findings may be interpreted to mean that when exploring a new situation, as in this experiment, the younger child spends less time on that exploration than does the older child, and a greater proportion of this exploratory time is tactual than is the case for the older child.

The lack of age trend on absolute tactual time may also be a function of the procedures used. While this study was intended to elicit active exploration in both visual and tactual modalities, the distinction between active and passive touch (Gibson, 1962) could not be adequately measured on the tests used. Schachtel (1959) has suggested a developmental shift from autocentric (sense-oriented) touch to allocentric (object-oriented) touch; while Ukhtomski (1961) has distinguished between defensive touch via pain receptors and active touch, via tactile receptors. If this distinction were tested in further research, an increase in age could reveal an increase in the proportion of active or allocentric touch, and a decrese in defensive or autocentric touch.

The expectation that schizophrenic children will express less visual preference than normal children of the same age was supported by the findings, thus also confirming Goldfarb's (1956, 1961) receptor hypothesis. There had been no age trend for absolute tactual time spent by normal children and there was no significant difference between schizophrenics and normals on these measures. However, the schizophrenic children may be using visual avoidance to ward off awareness of anxiety reduction in reliance on visual cues and a concomitant reduction in the level of perceptual field dependence (as indicated by the degree of error in RFT performance). On the basis of this hypothesis, and since it had been shown that alcoholics are extremely field dependent (Karp, Poster, & Goodman, 1963; Witkin, Karp, & Goodenough, 1959), 15 male alcoholics underwent 1 hour of modified SD (lying on a bed, wearing opaque goggles) following a pretest RFT. A matched control group was allowed an hour of normal activity following their pretest. On posttest measures, the experimental group showed a significant ($p < .01$) reduction in RFT errors, while the controls showed a nonsignificant change. The difference between the two groups' change scores was significant at $p < .05$.

On the basis of the evidence cited, and the results of the preliminary alcoholic study, it is hypothesized that brief SD may lead to a significant decrease in the level of perceptual field dependence in a normal sample.

METHOD

Subjects: Twenty-two females and 19 males, selected from introductory psychology classes, served as subjects. They were randomly assigned to experimental (9 males, 11 females) and control (10 males, 11 females) groups. Mean age for all subjects was 19 years.

Apparatus: A 36 × 36 × 1 inch square frame of aluminum strips, supported by a 36 × 1 inch horizontal brace, was coated on one side with luminous paint and mounted vertically on a 6-foot upright of 1.5-inch aluminum tubing by means of a center pivot that allowed the frame to rotate 360 degrees in the fronto-parallel plane. A 34 × 1 inch aluminum "rod" was also coated on one side with luminous paint and mounted on the upright so it could pivot within, but independently of, the frame. Mounted on the back of the upright was a fixed protractor, against which a movable pointer (attached to the back of the rod) marked off the degree of displacement from true vertical and horizontal. With the exception of the face of the rod and frame, all parts of the apparatus were painted flat black to reduce their visibility (apparatus modified after Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954).

For use in SD, a 5 × 7 foot cellotex-walled cubicle was outfitted with a standard bed. A Grason-Stadler white noise generator, Koss stereophonic headset, opaque goggles, cardboard cuffs, and heavy cotton work gloves were also used during SD. Polaroid (red) goggles were used for controlling dark adaptation. All studies were conducted in a relatively quiet, concrete-walled basement room.

Procedure. During the 10-minutes dark adaptation period, all subjects were instructed that they would be asked to "Give me your instructions for adjusting the rod to the vertical (horizontal) in relation to your own body or the walls of the room, regardless of how the frame may be tilted." While the subject stood 6 feet from the RFT apparatus, the experimenter adjusted the rod in 3–5 degrees increments as the subject directed. Twelve trials were administered: 3 horizontal and 3 vertical with the frame tilted 30 degrees right, and 3 horizontal and 3 vertical with the frame tilted 30 degrees left. An identical procedure was followed for RFT_2, the posttest.

Immediately following the RFT pretest (RFT_1), experimental subjects were told of the SD procedure, instructed to remain awake and to move around as little as possible, and fitted with the opaque goggles, gloves, cuffs, and headphones. The noise generator was turned to an intensity sufficient to mask all random noises in the room. A female assistant or experimenter was always in attendance to ensure the following of procedures. At the end of 1 hour, the subject was removed from the cubicle and questioned as to whether there had been any hallucinatory activity, and to determine if the subject had remained awake. Instructions were administered and RFT_2 was given immediately thereafter.

Following RFT_1, control subjects were led into the cubicle, instructed to remain awake, move around as little as possible, not get up, and leave their Polaroid dark adaptation goggles on. The subjects were allowed to read (with goggles on), and a radio was turned on sufficiently loud to mask any random noises in the room. The subject was always attended to ensure the following of procedures, and at the end of 1 hour was questioned about remaining awake. Immediately thereafter, the subject was removed from the cubicle, given RFT instructions, and the posttest.

RESULTS

Errors were scored as the sum of the deviation from true vertical and horizontal over the 12 trials. Because there were no significant differences between horizontal and vertical errors, the results were pooled.

The mean sum of errors on RFT_1 for control subjects was 44.6 degrees, and 60 degrees for experimental subjects ($t = 1.24$, *ns*). On the posttest the SD group's mean error was 43.95 degrees, a reduction of 16.05 degrees that was significant at $p < .01$, while control subjects showed a reduction of 5.56 degrees, a nonsignificant change. The difference between experimental and control groups' mean error reduction was significant at $p < .05$.

The data indicate that a brief period of SD serves to reduce perceptual field dependence on the RFT, and supports an hypothesis of increased saliency and availability of bodily cues.

DISCUSSION

The experimental group's significant reduction in error on the RFT after 1 hour of SD indicates support for an hypothesis of increased body awareness. Artificially modifying the sensory environment so that external stimulation is reduced may serve to increase the saliency and availability of internal cues (Zuckerman et al., 1962). The subsequent focusing of attention on internal stimuli is assumed to act to increase one's awareness of one's own body. That this does occur is indicated in several studies using quantitative measures of changes in body awareness (e.g., Azima et al., 1962; Reitman & Cleveland, 1964), as well as those involving verbal reports (e.g., Bexton et al., 1954; Freedman et al., 1961; Smith et al., 1961).

That learning or practice is little involved in the improved RFT performance has been demonstrated in the 3-year longitudinal studies by Witkin et al. (1962), who report test-retest reliabilities in the high .80's. Also, Elliott and McMichael (1963) have shown that RFT performance is not amenable to any consistent change through instruction and practice.

While it has been shown (Heron, 1961) that performance on the Embedded Figures Test (EFT), a measure highly correlated with the RFT, shows a decrement after SD, it must be noted that the EFT involves only the ability to overcome an embedding context or resist distraction. Although the RFT requires the same ability, there is also the opportunity to make use of bodily sensations in the task, a variable not involved in the EFT. Heron's (1961) report of spatial disorientation following SD appears incompatible with the present results, until one considers that his measure was a paper-and-pencil task.

The focusing of attention on, and increased awareness of, internal bodily sensations appears to reduce reliance on visual cues during the RFT. That is, field-dependent persons may become less so because the resolution of the conflict between visual and bodily cues inherent in the RFT situation is more heavily weighted on the side of the newly "recognized" and available internal sensations.

REFERENCES

ASCH, S., & WITKIN, H. Studies in space orientation: I. Perception of the upright with displaced visual fields. *Journal of Experimental Psychology,* 1948, 38, 191–196. (a)

ASCH, S., & WITKIN, H. Studies in space orientation: II. Perception of the upright with displaced visual fields and with body tilted. *Journal of Experimental Psychology,* 1948, 38, 217–223. (b)

AZIMA, H., LEMIEUX, M., & AZIMA, FERN. Isolement sensoriel: Etude psychopatholoique et psychoanalytique de la regression et du schema corporel. *L'evolutions Psychiatrique,* 1962, 2, 259–282.

BEXTON, W., HERON, W., & SCOTT, T. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology,* 1954, 8, 70–76.

ELLIOTT, R., & MCMICHAEL, R. Effects of specific training on frame dependence. *Perceptual & Motor Skills,* 1963, 17, 363–367.

FREEDMAN, S., GRUNEBAUM, H., & GREENBLATT, M. Perceptual and cognitive changes in sensory deprivation. In P. Solomon, P. Kubzansky, P. Leiderman, J. Mendelson, R. Trumbull & D. Wexler (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 59–71.

HERON, W. Cognitive and physiological effects of perceptual isolation. In P. Solomon, P. Kubzansky, P. Leiderman, J. Mendelson, R. Trumbull. & D. Wexler (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 7–33.

KARP, S., POSTER, DOROTHY, & GOODMAN, A. Differentiation in alcoholic women. *Journal of Personality,* 1963, 31, 386–394.

KUBIE, L. Theoretical aspects of sensory deprivation. In P. Solomon, P. Kubzansky, P. Leiderman, J. Mendelson, R. Trumbull, & D. Wexler (Eds.), *Sensory deprivation.* Cambridge: Harvard Univer. Press, 1961. Pp. 208–220.

REITMAN, E., & CLEVELAND, S. Changes in body image following sensory deprivation in schizophrenic and control groups. *Journal of Abnormal and Social Psychology,* 1964, 68, 168–176.

ROBERTSON, M. Theoretical implications of sensory deprivation. *Psychological Record,* 1961, 11, 33–42.

SMITH, S., THAKUROS, H., & LAWES, T. Perceptual isolation and schizophrenia. *British Journal of Psychiatry,* 1961, 107, 983–994.

WITKIN, H. The nature and importance of individual differences in perception. *Journal of Personality,* 1949, 18, 145–170.

WITKIN, H., & ASCH, S. Studies in space orientation: III. Perception of the upright in the absence of a visual field. *Journal of Experimental Psychology,* 1948, **38**, 603–614. (a)

WITKIN, H., & ASCH, S. Studies in space orientation: IV. Further experiments on perception of the upright with displaced visual fields. *Journal of Experimental Psychology,* 1948, **38**, 762–782. (b)

WITKIN, H., DYK, RUTH, FATERSON, HANNA, GOODENOUGH, D., & KARP, S. *Psychological differentiation.* New York: Wiley, 1962.

WITKIN, H., KARP, S., & GOODENOUGH, D. Dependence in alcoholics. *Quarterly Journal of Studies in Alcoholism,* 1959, **20**, 493–504.

WITKIN, H., LEWIS, H., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, P., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

YOUNG, H. A test of Witkin's field-dependence hypothesis. *Journal of Abnormal and Social Psychology,* 1959, **59**, 188–192.

ZUCKERMAN, M., ALBRIGHT, R., MARKS, C., & MILLER, G. Stress and hallucinatory effects of perceptual isolation and confinement. *Psychological Monographs,* 1962, **76**(30, Whole No. 549.)

(Received March 25, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 119–123

EFFECT OF FEAR AROUSAL UPON TASK PERFORMANCE AND VERBAL BEHAVIOR[1]

JAMES H. GEER[2]

University of Pennsylvania

80 undergraduate females were classified as being either high or low in fear of speaking in front of a group. Half were told that they were to speak following a task; the remaining Ss did the task without expecting to speak. All Ss spoke about the task for 1 min. immediately following the task. Experimental Ss took significantly longer to perform the task, and naive judges were able to detect differences in the verbal productions of the groups. Formal aspects of S's verbal productions also discriminated among high- and low-fear Ss. Implications for theory and importance of assessing different types of behavior are briefly discussed.

Anxiety or fear is purported to effect many different aspects of human behavior. The study reported in this paper attempts to assess the effects of fear arousal on two classes of behavior. Krause (1961) noted that performance on psychomotor tasks has been shown to be effected by arousal of anxiety. Mahl (1961) noted that certain formal aspects of speech are affected by emotional arousal; and Markel, Meisels, and Houch (1964) reported that naive judges were able to detect differences among the verbal productions of certain psychopathological groups even though they were not necessarily experiencing emotional arousal. This study was designed to extend the above findings to both performance on a psychomotor task and to the verbal production of "normal" subjects who are experiencing the experimental arousal of fear.

The general experimental design employed was as follows. The subjects were selected as being either high or low in the amount of fear they felt toward speaking in front of a group. One half of the subjects were told that they were to complete a task and were then immediately to speak extemporaneously about the task before a strange group. The other half of the subjects were told they had a task to complete and were not told that they were to speak following the task.

Immediately following the completion of the task, all subjects spoke for 1 minute, and these speeches were recorded. Experimental subjects were ostensibly speaking in front of a group; control subjects were simply recording their views of the just completed task. Tape recordings from approximately one half of the subjects noted above were played to a group of untrained judges who were uninformed as to the actual procedure. The judges were instructed to rate their estimation of the emotion the individual was experiencing.

METHOD

Subjects

All subjects used in this study were undergraduates from the Introductory Psychology Course at the State University of New York at Buffalo. Eighty females were employed in the initial portion of this study. Forty were designated as "high fear" (HF) and 40 as "low fear" (LF) on the basis of their response to the item "Speaking in front of a group" in the Fear Survey Schedule—II (Geer, in press). A subject was considered HF if she rated her fear of speaking in front of a group as either "Terror" or "Very Much"; the subject was considered LF if she rated the item as "None." The judges used in the second part of the study were 36 males and 24 females who were members of an evening section of Introductory Psychology.

Procedure

Twenty HF and 20 LF subjects were met by an experimenter and told that they were to participate in a study of extemporaneous speech. The experimenter conducted the subject to a room that contained a large one-way mirror, and the subject sat

---

[1] This research was supported in part by a grant from the Erie County Mental Health Foundation and in part by a grant from the Graduate School of the State University of New York at Buffalo, and was conducted while the author was at the State University of New York at Buffalo.

[2] The author is indebted to Sandra Mollenauer who collected the data of the study.

down facing the mirror. A tape recorder played through an amplifying system the following instructions to the subject:

> My name is Dr. Geer. I'm on the staff of the Psychology Department here at the University. I want to thank you for coming here today to participate in this study. Please listen carefully to what I say, as I can read the following instructions only once.
>
> We are studying the characteristics of individuals as they influence extemporaneous speaking before small groups. There are four people besides myself seated behind this one-way mirror. The mirror is built so we can see you but you cannot see us. We hope this will allow you to be a bit freer and more comfortable in your presentation while we are rating it. You'll be asked to talk extemporaneously for 5 minutes. We could have you speak on some familiar topic, but we thought it would be best if all participants had a similar experience, and then they could speak about that experience; therefore, you'll be asked to do a coding task. This is a common kind of psychological task and one that will give you material to talk about. As soon as you have finished the coding task, you will speak extemporaneously on your thoughts and experiences as they related to the task. The task is only to give you a common experience, so that you will have a background to speak on as the other subjects do. You will not be graded upon this task.
>
> Now I will have Miss Mollenauer describe the task to you. Again, let me thank you for your cooperation.

The subjects were not aware that the instructions were prerecorded and were led to believe that there were people behind the mirror.

Following the recorded message, the experimenter instructed the subject in the task to be performed. The task consisted of having the subject complete a page of the WAIS digit-symbol subtest. The subject was reminded that the extemporaneous speech would begin as soon as the page was completed. The experimenter sat behind the subject and recorded as many individual codings as were completed every 30 seconds, and the experimenter also recorded the total time each subject required. The recording of rate of coding and total time was done without the subject's knowledge.

When the subject finished the coding task, she was asked to stand in front of the one-way mirror and speak for 5 minutes on the task that she had just completed. The experimenter started the tape recorder and told the subject to begin. The subject was stopped after 1 minute. Following the taping, the subject was told that the recording would be used anonymously and that there had been no one behind the mirror. The subject was excused after the experimenter extracted a promise from her not to speak of the experiment until the end of the semester.

Control subjects were told that the experimenter was collecting normative data on the coding task. They were not told that they would speak nor were they told that there were any observers. When Control subjects finished the coding task, they were asked to speak into a tape recorder giving their views concerning the task they had just completed.

The second phase of this study consisted of having naive judges rate the verbal productions of the subjects employed in the first part of the study. The tape recordings of the experimental and control subjects were examined and recordings that were not clear or that gave any clue concerning the experimental manipulation were discarded.[3] From the remaining recordings, the subjects were randomly discarded until 40 recordings, 10 from each of the four subgroups (high and low fear, divided into experimental and control), remained. For each of the 40 selected recordings, the first 30 seconds of the speech was put onto a master tape with 15 seconds of silence between each 30 seconds of speech.

The master tape was played to the judges who were told that they were to rate the emotion they felt that the individual who made the recording was experiencing. Judgments were made on the 7-point rating scales presented below:

Estimated Subject's Reaction

Pleasure	1	2	3	4	5	6	7	Anger
Bored	1	2	3	4	5	6	7	Interested
Accepting	1	2	3	4	5	6	7	Rejecting
Tense	1	2	3	4	5	6	7	Calm

The ratings were made during the 15-second silent period between the speeches. Procedural errors made during the judging made two speeches unusable. To provide equal *N*s in the subgroups, two other subjects were randomly discarded to provide ratings on 36 subjects, 9 from each subgroup. Thus, 60 judges rated each of the 36 speeches on each of the above 4 dimensions.

Finally, the 36 recordings were examined for the speech-disturbance indexes reported by Mahl (1961). Mahl reports success in detecting transitory anxiety through the use of his "Speech Disruption Ratio (non-ah)" and the "Silence Quotient." These measures were obtained for each of the 36 recordings. Also, a rate of speech measure was obtained. This was the number of words emitted during the 30 seconds of recording.

RESULTS

The 2 × 2 factorial analyses of variance were the basic statistical tests used in this study. The factors were fear level and experimental conditions. The first such analysis was performed upon the total time to complete the coding task. A significant main effect of conditions ($F = 9.43$, $df\ 1/76$, $p < .01$) was found. Inspection of the data clearly revealed that experimental subjects took longer to complete the coding task than control subjects. The mean time to complete the coding task was 2 minutes and 5 seconds for experimental subjects and 1 minute and 41 seconds for control subjects. There were no significant differences between HF and LF subjects within the experimental condition.

---

[3] Judgments concerning clarity and presence of clues relating to experimental manipulation were made without knowledge of the subject's group assignment.

An analysis of the number of coding operations performed during successive 30-second periods was performed. There was no evidence that experimental subjects slowed more than control subjects as the end of the task approached. Also, there was no evidence that HF subjects differed from LF subjects in the differential rate of coding as the end of the task approached.

The ratings of subjects' verbal productions were analyzed. The dependent variable was the mean rating of each subject's voice. The 2×2 factorial analysis of variance of the "tense-calm" scale did not yield any significant results. Thus, naive judges were unable to detect voice differences among the subgroups when rating on a tense-calm dimension.

The analysis of variance performed on the "bored-interested" dimension yielded a significant ($F = 4.5$, $df\ 1/32$, $p < .05$) result. Judges rated HF subjects as less interested in the task than LF subjects.

The 2×2 analysis of variance which was performed upon the "pleasure-anger" dimension yielded a significant Fear Level × Conditions interaction ($F = 4.64$, $df\ 1/32$, $p < .05$). Inspection of the data suggested that the result was caused by HF subjects in the experimental condition being rated as more angry than all other groups. This was confirmed by applying the Newman-Keuls a posteriori test (Winer, 1962) which revealed significant differences between HF experimental subjects and all other groups ($p < .05$) but no significant differences among the remaining three groups.

The results of the analysis of the "accepting-rejecting" dimension were similar to the pleasure-anger findings. A significant Fear Level × Conditions interaction ($F = 4.64$, $df\ 1/32$, $p < .05$) was found. Verbal productions of HF experimental subjects were rated as indicating greater rejection than all other groups. This also was confirmed by the Newman-Keuls test which revealed a significant difference between HF experimental subjects and all other groups while no significant differences were found among the remaining three groups.

These data clearly reveal that naive judges can, on certain rating scales, discriminate among the groups and conditions. HF subjects under experimental conditions yielded verbal productions that were rated as more angry and more rejecting than other subjects. Also, HF subjects in general were rated as finding the experiment less interesting than LF subjects.

The final analyses relate to the formal characteristics of the verbal productions. Mahl's (1961) "Speech Disruption Ratio (non-ah)" was analyzed by the 2×2 factorial analysis of variance. The results did not attain accepted levels of statistical significance; however, the fear level F was significant at the 10% level ($F = 3.72$, $df\ 1/32$). The data suggested that HF subjects yielded more speech disruptions. The analysis of variance of the "Silent Quotient" did yield statistically significant results. The verbal productions of HF subjects had more silence than LF subjects ($F = 10.75$, $df\ 1/32$, $p < .05$). The mean silence time for HF subjects was 8.8 seconds and was 4.4 seconds for LF subjects.

The rate of speech (number of words in the 30 seconds) was analyzed and also yielded statistically significant results; HF subjects spoke more slowly than LF subjects ($F = 13.75$, $df\ 1/32$, $p < .01$). These measures indicate that the verbal productions of HF subjects have different formal qualities than the verbal productions of LF subjects. Further, it may be noted that these differences are consistent with a picture of increased tension or fear.

DISCUSSION

The results from the coding task are clear. The subjects who anticipated having to speak in front of a group showed lowered efficiency on the task as measured by time required to complete the task. The findings did not support the expectation that HF subjects in the experimental condition would show the greatest inefficiency. The most plausible explanation is that all experimental subjects were aroused by anticipating having to speak before a strange group and the task was not sensitive enough to detect differences between HF and LF groups.

Two theoretical positions may be used to explain the above results. The first, an interference theory, would suggest that the fear aroused by the experimental manipulation interfered with task performance. This interference led to an increase in the time required to finish the task. However, one piece of supplementary data does not support this position. It might be expected, from interference theory, that subjects under arousing conditions would yield more errors on the coding task. This was evaluated and there were no significant differences among groups in the number of coding errors.

A second theoretical position that might explain the data from the coding task would be Miller's (1959) conflict model. It can be assumed that experimental subjects viewed the anticipated speech as an unpleasant event. The coding task may be conceived as the path to this attempt. In that situation, the subject may attempt to delay the onset of the unpleasant event by slowing his response rate on the coding task. This, of course, describes what did occur on the coding task. As with the interference theory, a supplementary analysis may be employed in an attempt to verify the conflict model interpretation. Miller's conflict model states that the tendency to avoid a negative goal will increase the closer one approaches the goal. Thus, it would be expected that subjects in the experimental condition would show progressively reduced rates of responding over 30-second periods. As noted in the Results section, this expectation was not confirmed. Further experimentation will be necessary in order to account accurately for the findings relating to the coding task.

The results of rating the verbal productions of subjects for the "estimated subjects reaction" are of considerable interest. First, naive judges were not successful in discriminating among subject groups when rating verbal productions on a tense-calm dimension. The other rated dimensions did reveal that naive judges could discriminate successfully among verbal productions of subject groups. HF subjects' speech segments were rated as indicating that they were bored with the procedure when compared with those of LF subjects. Further, the verbal productions of HF subjects under experimental conditions were rated as indicating more anger and more rejection than all other groups. These results suggest that untrained judges are able to detect differences in the verbal productions of subjects who report fear of speaking in front of groups when the subjects believe they are speaking to a group.

Several points concerning the above findings require further discussion. First, verbal productions were a more sensitive index of the hypothesized condition of fear than performance on the coding task. Certainly, clinical psychologists would agree that it is possible to infer the emotional state of individuals from their verbalizations. This study lends some support to that position. An interesting point is that the detection was done by naive or untrained judges. It should be noted, however, that the judges did not accurately detect the hypothesized emotional state of fear. It would be expected that accurate detection of emotion from subjects' verbal productions would have resulted in the finding of significant differences on the tense-calm dimension. The reason for the discrepancy is not clear. It may be that the experimental condition only angered the subjects; however, the manipulation affected primarily those who were HF subjects. Possibly the naive judges could only detect emotional arousal and were not able to discriminate among emotions without awareness of the context of the emotional arousal.

The verbal productions of subjects in this study were not controlled and therefore it is not possible to determine accurately what aspects of subjects' speech were being responded to by the judges. The results of the analyses of the formal characteristics of the verbal productions yield some suggestions. The rate of speech was reduced and the length of silences increased for HF subjects. These two measures are obviously related;

as silence increases, the number of words that can be uttered must decrease. This was confirmed by computing the correlation between the rate and silence measures. The correlation was $r = -.81$, 35 *df*, $p < .01$. Other formal characteristics of speech not evaluated in this study may indicate differences among subjects experiencing emotional arousal.

It is quite possible that the content of subject's productions is more important than the formal characteristics of the speech. Judges rating on two dimensions were able to discriminate HF and LF subjects under experimental conditions. Formal speech characteristics distinguished among HF and LF in general but not fear level in interaction with the experimental conditions. These results indicate that in this study judgments more accurately discriminate among groups than formal characteristics of the voice. This suggests that perhaps the content of the verbalizations, which may be detected by judges even when obvious clues as to experimental condition are excluded, is the important variable.

The results of the study clearly support Krause's (1961) suggestion that when studying emotions, one should, where possible, look at all aspects of behavior. Data of interest were obtained from both the coding task and from the verbal productions of the subjects. Exclusion of either type of data would have reduced the meaningfulness of the study. This study confirms previous work in this laboratory (Geer, in press) which indicated clear-cut behavioral differences between HF and LF subjects as selected by the FSS-II.

REFERENCES

GEER, J. H. The development of a scale to measure fear. *Behaviour Research and Therapy,* in press.

KRAUSE, M. S. The measurement of transitory anxiety. *Psychological Review,* 1961, **68**, 178–189.

MAHL, G. F. Measures of two expressive aspects of a patient's speech in two psychotherapeutic interviews. In L. A. Gottschalk (Ed.), *Comparative psycholinguistic analysis of two psychotherapeutic interviews.* New York: International Univer. Press, 1961. Pp. 91–114.

MARKEL, N. N., MEISELS, M., & HOUCK, J. E. Judging personality from voice quality. *Journal of Abnormal and Social Psychology,* 1964, **69**, 458–463.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology, a study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received March 29, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 124–135

TREATMENT OF ANXIETY THROUGH SYSTEMATIC DESENSITIZATION IN THERAPY GROUPS[1]

GORDON L. PAUL AND DONALD T. SHANNON

University of Illinois

The effects of short-term treatment by modified systematic desensitization in time-limited intensive therapy groups were evaluated in a matched groups design. 10 chronically anxious college males, treated by the group method, were evaluated on the basis of personality and anxiety scales against an "own-control" period, and 4 equated groups of 10 Ss each. 1 group served as an untreated control for evaluating extra-treatment effects on college grade-point average (GPA), as an objective, public criterion. Ss in the remaining groups received 1 of the following individual treatments: systematic desensitization, insight-oriented psychotherapy, or an attention-placebo treatment. The group method produced several significant improvements, suggesting that combined group desensitization offers an efficient and effective treatment for social-evaluative anxiety.

The shortage of therapists, increased demands for services, and questionable efficiency and effectiveness of long-term individual psychological treatment in the traditional mode have long been recognized (Schofield, 1964). The most promising attempts to extend existing services have come through the introduction of group treatment techniques (Hunt, 1964), time limited short-term treatment (Frank, 1959; Muench, 1964; Schlien, Mosak, & Driekers, 1960), and development of treatment procedures based upon learning principles (Bandura, 1961). Lazarus (1961) has reported a double economy of effort by treating individuals with phobic disorders in groups of two to four with systematic desensitization. Unfortunately, empirical data on the utility of specific treatment innovations in groups is still quite limited.

In a recent controlled investigation of comparative outcome effects of short-term individual treatment (Paul, 1965), a modified form of Wolpe's (1961) systematic desensitization was found to be significantly more effective in reducing maladaptive anxiety than were other forms of treatment. In Paul's study, the public speaking situation was taken as a prototypic stress condition for the elicitation of interpersonal performance anxiety. Ninety-six subjects, most debilitated by anxiety, were selected from students who requested treatment following administration of a battery of personality and anxiety scales to 710 public-speaking students. Cognitive, physiological, and observable behavioral measures of anxiety under stress conditions all revealed a significantly greater reduction in anxiety, from pre- to posttreatment, for subjects treated by a modified desensitization procedure than for equated groups who received insight-oriented psychotherapy or attention-placebo treatment administered by the same experienced therapists. Significant gains were also obtained relative to an untreated "wait-list" control group, and an untreated "no-contact" control group. The greater anxiety reduction achieved through systematic desensitization was also evident in therapist's ratings, and was maintained at a 6-week follow-up with no evidence of "symptom substitution."

Although the primary result of Paul's (1965) study was empirical support for a learning approach to interpretation and treatment of emotional disturbances, another significant aspect was the short-term treatment involved—only five sessions. While treatment focus and effects were mainly limited to anxiety experienced in speaking before groups,

---

[1] This study was partially supported by Public Health Services Fellowship 1 F1 MH-19, 873, 01 from the National Institute of Mental Health, held by the senior author. Data reported here for comparison purposes were drawn from an earlier study supported in part through the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare.

generalized effects of the desensitization treatments were reported by several subjects. Unfortunately, the timing of the study prohibited extra-experimental measures of generalized effects. Additionally, it seems possible that the public speaking course itself enhanced the effects due to treatment, once a degree of anxiety reduction had been attained (Paul, 1965).

Recent studies (Spielberger, 1962; Spielberger & Katzenmeyer, 1959) have reported that highly anxious students receive lower grades and a higher academic failure rate than nonanxious students of equal intelligence. Paul and Eriksen (1964) also found significant detrimental effects on course examinations for subjects high on examination anxiety when ability level was taken into account. Since the primary bases of grading at the university level are examinations and classroom participation, the change in grade-point average (GPA) provides an excellent external measure of effectiveness in the treatment of interpersonal performance anxiety.

Due to ineffectiveness of treatment, or confounding factors, the results of psychological treatment for college students have failed to produce significant changes in GPA (Spielberger & Weitz, 1962). Spielberger and Weitz report improved GPA from midterm to semester's end for a group of anxious students treated by group methods. However, their results are difficult to interpret due to the presence of a unique personality pattern in a subgroup of nine of their 26 clients which accounted for the overall improvement.

In the present study, we have attempted to extend Lazarus' (1961) group innovation to include social-evaluative phenomena in the *short-term* treatment of interpersonal performance anxiety by systematic desensitization in the context of intensive group therapy. The major purpose was to evaluate the feasibility and effectiveness of administering modified systematic desensitization in groups, combined with re-educative discussion, by (*a*) evaluating changes on personality and anxiety scales following combined group desensitization against a comparable untreated wait-period in "chronically" anxious clients; (*b*) evaluating the effects of combined group desensitization on an extra-experimental index of performance, the GPA, against a matched untreated control group; (*c*) comparing pre-post changes resulting from the group method with those obtained previously for comparable subjects with individual treatments, in order to evaluate the extent of effectiveness of the group procedure.

METHOD

Subjects

A total of 50 highly anxious male subjects, ranging in age from 19 to 24 years (mdn = 21) participated in the present study. All were undergraduates, (mdn = Soph) enrolled in a required public speaking course during the fall semester at the University of Illinois. All subjects completed a battery of personality and anxiety scales at the beginning of the semester following their first classroom speech. The battery included forms of (*a*) IPAT Anxiety Scale Questionnaire (Cattell, 1957); (*b*) Pittsburg Social Extroversion-Introversion & Emotionality Scales, including the MMPI L-scale (Bendig, 1962); (*c*) Interpersonal Anxiety Scales (Speech before a large group, Competitive contest, Job interview, Final course examination) of the S-R Inventory of Anxiousness (Endler, Hunt, & Rosenstein, 1962); (*d*) Personal Report of Confidence as a Speaker, PRCS short-form (Paul, 1965).[2] Each subject was selected from a population of 710 students who completed the initial battery on the basis of indicated motivation for treatment, highest scores on performance anxiety scales (PRCS ≥ 16; SR-speech ≥ 35) and low falsification ($L < 7$).

Procedure

The basic plan of the study is presented in Table 1. It consisted of five groups of 10 subjects each, all of whom are drawn from the larger groups participating in the earlier study (Paul, 1965) during the first semester. The five groups were equated (*M* and *SD*) on performance anxiety scales of the test battery obtained on Test 1. The three groups receiving individual treatment (systematic desensitization, insight-oriented psychotherapy, or attention-placebo) were treated during the first semester with concurrent enrollment in the speech course. The 10 subjects assigned to the combined group desensitization procedure were treated during the second semester, after completion of the speech course. These latter subjects were selected from the "wait-list" controls of the earlier study who showed *no* reduction in anxiety scores over the wait period, in addition to matching individual treatment groups on performance anxiety measures. The untreated controls were drawn from "uncontacted" subjects of the first se-

---

[2] The original battery also included a form of the Anxiety Differential (Husek & Alexander, 1963). This form was excluded from analysis for the present investigation since a stress administration was not included.

TABLE 1

EXPERIMENTAL DESIGN SHOWING THE TIMES AT WHICH SUBJECTS WERE EVALUATED (TEST BATTERY, GPA)

Group	Experimental procedures				
	First Semester		Second Semester		
Group desensitization Untreated controls	Test 1 (Pre$_1$) Test 1	Wait period —	GPA Test 2 (Pre$_2$) Test 2 GPA	Treatment —	GPA Test 3 (Post) GPA
Systematic desensitization Insight-oriented psychotherapy Attention-placebo	Test 1 (Pre) Test 1 (Pre) Test 1 (Pre)	Treatment Treatment Treatment	Test 2 (Post) Test 2 (Post) Test 2 (Post)	Individual treatments	

mester for comparison of effects of group desensitization on GPA. The latter subjects were equated with group desensitization subjects on college, class, and age as well as on *all* scales of the test battery ($p > .10$ for all t and F tests).

After classroom administration of the test battery to the speech course population (Test 1) all subjects except untreated controls were contacted individually by the senior author and seen in a short (15-20 minutes) interview for screening and instructions. Common expectations with regard to treatment were instituted by informing each subject that he fell among the top 10-15% on performance anxiety measures, and briefly presenting the rationale and course of treatment to be undergone. The subjects later assigned to group desensitization were told that their names had been "picked from a hat," and that they could not be treated during the first semester due to a shortage of therapists; however, that treatment would be provided the second semester. Those subjects assigned to the untreated control group were never contacted individually and were unaware of their participation in a therapy study.

Four weeks after Test 1, the individual treatments were begun. These three treatments were conducted concurrently for 5 hours over a 6-week period, with missed appointments rescheduled during the week. The subjects in all five groups continued in the speech course during the first semester. Specifics of subject contact, communications, and additional measures obtained during the first semester are reported in detail in Paul's monograph.

Six weeks following termination of individual treatments, the test battery was again administered to the entire population (Test 2), and improvement ratings given by treated subjects. Test 1 and Test 2 thus provided for evaluations pretreatment and at a 6-week follow-up for individual treatment subjects. These two testing periods delineate the "wait period" for subjects later treated by group desensitization.

Following the second administration of the test battery (Test 2), the "chronically anxious" subjects were selected from the untreated controls of the first semester according to the criteria described above. Those assigned to the combined group desensitization treatment were further subdivided into two treatment groups[3] (based on within-group homogeneity of anxiety scores and age, splitting extroversion scores as the restricted range allowed) and interviewed by their prospective therapists. As with individual treatments, this initial interview served the purpose of instituting common expectations and presenting the course of treatment to be undergone.

Combined group desensitization was started 4 weeks after Test 2 and continued for a total of nine sessions on a weekly basis. Any subject who missed two successive group meetings was brought in for a make-up session to maintain continuity of completed hierarchy items. The test battery, including self-ratings of improvement, was again completed (Test 3) and mailed in by subjects in this treatment group 4 to 6 weeks after treatment termination (after semester's end).

After Test 2, the subjects treated individually and untreated controls were not contacted further, but names of the matched untreated controls were placed on file for later determination of academic performance. Final GPA for first (wait) and second (treatment or control) semesters for the group desensitization and untreated control groups were then obtained from the University Registrar during the summer following the second semester.

Thus, the subjects treated by combined group desensitization constiute an "own-control," plus allowing comparison with a matched "untreated" group on GPA, and comparisons over equal time intervals with the earlier individually treated groups on personality and anxiety scales.

Therapists

Five clinical psychologists, all highly experienced psychotherapists (6–18 years), administered the individual treatments. Each therapist treated two subjects from all three individual treatment groups. The five therapists (one of whom was the junior author) were trained in the administration of individual systematic desensitization and attention-placebo treatments by the senior author. All treatment contacts were monitored and supervised via tape recordings and weekly meetings. Responses to 24 attitudinal and 20 technique scales on the Therapist Orientation Sheet (TOS) revealed these therapists to be representative of recent national samples, and to differ only slightly among themselves (Paul, 1965). Analyses of therapist effects in the earlier study, with larger groups, found no significant differences between therapists in overall or differential effectiveness.

---

[3] A third group of five subjects had been planned initially, but were excluded after failure to obtain complete data on four of the five subjects, and timing difficulties which prevented total attendance at any group meeting. Appreciation is expressed to Thomas J. D'Zurilla for serving as therapist for this group and participation in planning and progress sessions.

The present authors served as therapists for the combined group desensitization treatment. Both were clinical psychologists with 3 and 12 years of clinical experience, respectively. Prior to participation in the present study, relevant experience included 30 individual systematic desensitization cases and 3 intensive groups for one therapist, and 6 intensive groups and 12 individual desensitization cases for the other. Both therapists considered their primary "school" identification to be eclectic, with secondary affiliation indicated to be neobehavioristic, for one, and Sullivanian for the other. For both therapists, two writers indicated to be most influential in shaping approach to treatment were A. Lazarus and J. Frank. G. Kelley and J. Wolpe were also considered influential.

Responses of the group therapists on the TOS revealed them to be quite similar to the individual therapists, and remarkably similar to each other.[4] The two therapists did not differ by more than one scale point, of the five possible, on the 24 attitudinal scales, and differed on only two of the 20 technique scales: use of reassurance and approval. Comparability of the two desensitization groups was further insured by weekly discussions of procedure and progress.

Individual Treatments

The individual treatments are explained in detail elsewhere (Paul, 1965), complete with procedural manuals for modified systematic desensitization and attention-placebo treatments. Therefore, only a brief resume of these treatments will be presented here. All individual treatments were time-limited to five contact hours over a 6-week period. The same five therapists administered all three treatments, thus holding constant across groups the potentially important effects of therapist prestige, personality, and physical attributes, as well as time and attention received. These treatments were administered equally at the University Psychological Clinic, Student Counseling Bureau, and Guidance and Counseling Center. Each subject was again presented with a "rationale" for the treatment to be undertaken during the first treatment session.

Insight-oriented psychotherapy consisted of the usual interview approach utilized by the respective therapists in their daily work. In this treatment, anxiety reduction was attempted through helping the client to gain "insight" and "self-understanding" of the historical and current bases and interrelationships of his problem. The focus was with the public-speaking situation, and all therapists indicated a high degree of confidence in effecting therapeutic change for the specific situations within the time allowed.

Systematic desensitization consisted of a modified and formalized form of the treatment advanced by Wolpe (1961). The major features of this treatment included: (*a*) training subjects in deep muscle relaxation by a much abbreviated form of the Jacobsen technique; (*b*) constructing a spatial-temporal "anxiety hierarchy" related to public speaking (a graded series of events falling along a stimulus generalization gradient of distance outward in time and space from the most threatening public performance), and; (*c*) attempting to counter-condition "items" (events) from the hierarchy by directing subjects to image the events while in a state of deep muscle relaxation. The images proceeded from least to most threatening items according to specific rules of timing. The hierarchies were specifically constructed to a public speaking situation.

The attention-placebo treatment was included to determine the degree of improvement which might be attributed to nonspecific treatment effects, such as: expectation of relief; attention, warmth, and interest of the therapist (therapeutic relationship); suggestion, and "faith." The subjects receiving this treatment were told that it was a method of training to work and think effectively under stress. A "fast acting tranquilizer" (5 grams capsule of sodium bicarbonate) was administered at the beginning of each session, with pulse rate and pupilary response taken after a period of silence. The subject then undertook a "very stressful task" (identifying nonnoxious sounds from a tape recording) while the therapist watched attentively, recording the client's "responses." The subject was told that anxiety would not be experienced during this task because of the effects of the "tranquilizer." At the end of each session subjects were assured that treatment was progressing well. This proved to be an excellent control treatment since no verbal interaction could transpire during 45 minutes of each session, and subjects easily accepted the "rationale."

Combined Group Desensitization

The experimental treatment of focal interest consisted of the group application of the individual modified systematic desensitization plus intensive group discussion with re-educative goals. This treatment was conducted in closed-groups of five clients each, time-limited to nine sessions on a once-per-week basis. All sessions were conducted at the University Psychological Clinic in a therapy room with dimensions of 9 × 12 feet. Five overstuffed leather chairs were arranged in a semi-circle, facing the therapist, and the room was furnished with carpet and drapes to minimize external stimuli.

---

[4] Mimeographed copies of responses to the Therapist Orientation Sheet by both group therapists may be obtained by writing either author, University of Illinois, Psychological Clinic, 1005 W. Nevada, Urbana, Illinois 61801.

A schedule was provided each therapist which gave the amount of time to be spent in various activities during each successive treatment hour, and each session was recorded on tape to insure comparability.

The basic procedures followed by both therapists in the discussion periods were aimed at increasing confidence, skills, and awareness of effects in interpersonal relationships (discrimination learning). The major group techniques utilized included: generalizing from one client's remarks; emphasizing similarities between problems, experiences, emotions, etc.; referring questions to the group; applying general statements to the immediate group situation; and pointing out difficulties in communication.

During the first treatment hour a maximum of 10 minutes was devoted to personal introductions and a statement of degree, duration, and extent of social or evaluative anxiety by each client. Five to 10 minutes was spent in presentation of rationale and course of treatment. The rationale presented was limited to a brief statement that subjects' emotional reactions were the result of previous experience, and that these inappropriate reactions could be unlearned. This could be accomplished by determining the situations in which they become anxious, constructing a hierarchy of situations from least to most anxiety producing, and repeatedly visualizing these situations while deeply relaxed. It was also explained that the advantage of relaxation was that they could not be both tense and relaxed at the same time. Further, in gradually proceeding up the hierarchy, through imagery, the relaxed state would desensitize the previously anxiety provoking situations.

Since all subjects reported the speech situation to be the most anxiety provoking of all interpersonal situations, treatment was first directed towards desensitization of speaking situations. This decision was made on the assumption that other less threatening interpersonal experiences would also be diminished via generalization. Clients were told that specific desensitization of other areas of anxiety would be carried out later in treatment if desired.

The following 30–35 minutes of the first treatment hour consisted of training in progressive relaxation following the accelerated procedure and rules of timing presented by Paul (1965), with appropriate modifications dictated by group rather than individual training.[5]

After completion of relaxation training, subjects were aroused and the remaining 5 to 10 minutes of the session devoted to a discussion of feelings and problems experienced during relaxation training. Clients were then instructed to practice relaxation for 15 minutes twice per day between sessions, focusing on muscle groups which had not responded during the initial hour.

The first 10 minutes of the second session were spent discussing the success of relaxation practice and correcting problems or misconceptions involved with the relaxation training. The next 10–15 minutes consisted of construction of the first hierarchy by group discussion. This was a spatial-temporal anxiety hierarchy relative to public speaking. The final speech from the previous semester's course was taken as a common reference point, extending the hierarchy into the future with situations involving larger and more threatening audiences. Hierarchies were constructed to include the common situational elements of all clients in the group, with the items geared to the most anxious member. However, items were worded to allow each subject to include the unique aspects of his own situational experiences. The actual hierarchy for one group contained 16 items (extended to 18 during the course of treatment) from, "You are discussing your approaching speech with friends two weeks before it is due," as the lowest item, to "You are standing in the center of the Assembly Hall presenting a speech to a packed house," as the highest. The hierarchy for the second group contained 14 items, from "You are beginning to do background reading, two weeks before the day of the speech," to, "You are standing, facing a packed audience in an auditorium, ready to speak."

The next 20–25 minutes of the second session consisted of additional training in progressive relaxation followed by a test of imagery. The latter involves asking subjects to image a common, nonanxious scene and evaluating its formal aspects, as well as determining whether anxious scenes did in fact arouse anxiety for all subjects. Each subject's facility with relaxation and imagery was then discussed and problems corrected in the remaining meeting time.

Construction of the initial hierarchy was completed during the first 15 minutes of the third session, and discussion of problems related to the use of relaxation again handled during this period. Relaxation was induced, as previously, but requiring only 15–20 minutes. The remaining time (to the last 5 minutes) was devoted to presentation of images from the hierarchy, following the formal rules of timing presented by Paul (1965), gearing movement to the "slowest" group member.[6] During the last 5 minutes of the third session reactions and problems with relaxation and desensitization procedures were discussed. Clients were instructed to stop practicing relaxation between sessions, but to use it in stressful situations.

---

[5] Briefly, this consists of dimming the lights and instructing clients to close their eyes and sit in such a manner that their bodies are entirely supported by their chairs. Clients are then instructed to successively tense and hold gross muscle groups throughout the body, focusing attention only on the specific muscle group of the moment, until the therapist says "relax" (7 seconds), and then to *immediately* release the muscles and attend only to the relaxation which takes place (20 seconds). The first muscle group consists of the dominant hand and forearm, followed by upper arm, nondominant hand and forearm, upper arm, forehead, facial muscles, etc., throughout the body. In the first training session a minimum of two tension-release cycles are covered for each muscle group. In each instance the therapist proceeds to the next muscle group when the first is reported to be completely tension-free, proceeding at the pace of the "slowest" client in the group. Since talking during this phase is to be kept at a minimum, questions by the group therapist are phrased in the negative, with subjects' answers by hand signals, for example, "if your abdominal muscles are not as relaxed as your chest muscles, signal by raising your right hand."

Sessions 4 through 8 were conducted according to the following time schedule: the first 15 minutes were devoted to discussion of problem areas, including changes in anxiety hierarchies or construction of new hierarchies; relaxation was induced, requiring 8–15 minutes; and, remaining time was devoted to imaging items from the hierarchies, with the last 5 minutes of each session devoted to a discussion of reactions during desensitization.

The induction of relaxation required less time with each session as subjects became more practiced and achieved greater discrimination. For Sessions 4 and 5, tension-release cycles were directed at even larger muscle groups, e.g., both hands and arms together. By Session 6, all subjects were able to achieve deep relaxation by focusing attention on gross muscle groups and imaging the previous experiences of release without first tensing muscles. Although such a procedure bears an obvious similarity to hypnotic training, no mention of "hypnosis" was made, and therapists actively avoided any attempts at direct suggestion.

During Session 7, a test anxiety hierarchy was constructed for both groups, involving seven items for one group and nine for the other. As with the speech hierarchy, the test hierarchy was a spatial-temporal stimulus generalization gradient directed at the most threatening course examination for each subject.

The ninth and final session was included to allow completion of anxiety hierarchies, should this have proven necessary, and also to provide closure through a discussion of treatment effects, future plans, and goals. For one group, the speech hierarchy and test hierarchy were both completed during the eighth session so that the entire final session was spent in discussion. For the other group, the speech hierarchy was completed during the eighth session, but about 10 minutes of the final session were needed to complete three items from the test hierarchy.

RESULTS

Test Battery Data over Wait and Treatment Periods for Group Desensitization Ss

Of first concern was the evaluation of changes between wait and treatment periods for subjects treated by combined group desensitization, and a test of differences attributable to the two therapists administering this treatment. A three-way trend analysis (Therapists, Pre_1-Pre_2-Post, Subjects) was performed on data for each scale of the test battery. The summary of these analyses for the group subjects is presented in Table 2 for specific anxiety scales, and Table 3 for general scales. The means and *SD*s for all

TABLE 2

ANALYSES OF VARIANCE ON SPECIFIC ANXIETY SCALES OVER "WAIT" AND TREATMENT PERIODS FOR *S*s TREATED BY GROUP DESENSITIZATION ($N = 10$)

Source	*df*	PRCS		SR-Speech		SR-Exam		SR-Interview		SR-Contest	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Therapists	1	97.20	4.18	218.70	6.79*	282.13	1.12	3.34	—	67.50	—
a. error	8	23.28		32.20		251.97		136.18		142.18	
Pre_1-Pre_2-Post	2	304.03	31.64**	525.55	38.01**	133.30	7.31**	221.44	4.14*	133.44	5.05*
linear component	(1)	451.25	46.97**	451.25	32.64**	140.45	7.70*	174.05	3.25[a]	174.05	6.69*
quadratic component	(1)	156.82	16.32**	599.85	43.39**	126.15	6.92*	268.82	5.02*	88.82	3.41[a]
Therapist × Pre_1-Pre_2-Post	2	17.10	1.78	35.35	2.56	.44	—	5.63	—	.30	—
b. error	16	9.61		13.83		18.24		53.53		26.03	

[a] $p < .10$.
* $p < .05$.
** $p < .01$.

---

[6] The only change in item presentation, from individual desensitization, was that each desensitization session began with a presentation of all new items covered in the previous session. This was instituted in order to avoid the necessity of "make-up" sessions should any subject be unable to attend a single meeting.

TABLE 3

ANALYSES OF VARIANCE ON GENERAL SCALES OVER "WAIT" AND TREATMENT PERIODS FOR SUBJECTS TREATED BY GROUP DESENSITIZATION ($N = 10$)

Source	*df*	Scale					
		Extroversion-Introversion		Emotionality		IPAT Anxiety	
		MS	*F*	*MS*	*F*	*MS*	*F*
Therapists	1	128.14	—	70.54	—	104.54	—
a. error	8	181.97		81.10		347.08	
Pre_1-Pre_2-Post	2	133.64	17.32**	36.44	3.71*	163.04	4.62*
linear component	(1)	266.45	35.53**	64.80	6.60*	211.25	5.99*
quadratic component	(1)	.82	—	8.07	—	114.82	3.26[a]
Therapist × Pre_1-Pre_2-Post	2	11.63	1.51	1.63	—	8.23	—
b. error	16	7.72		9.83		35.26	

[a] $p < .10$.
* $p < .05$.
** $p < .01$.

scales under the three administrations are presented in Table 4.

Inspection of Tables 2 and 3 reveals no significant main effect for therapists except on the SR-speech scale, due to higher initial scores on that scale for one treatment subgroup. More importantly, the therapist by Pre_1-Pre_2-Post interaction failed to approach significance on a single scale. The latter finding indicates that the subgroups did not differ in movement over time, and further, that changes effected by the two therapists were comparable. Additionally, each therapist received equal mean ratings of "Very Competent" and "Very Likeable" from subjects in their respective groups. Therefore, data for the two subgroups receiving group desensitization are combined for further analyses.

TABLE 4

MEAN SCORES ON SELF-REPORT SCALES OVER TWO PRETESTS (WAIT PERIOD) AND POSTTEST (TREATMENT PERIOD) FOR SUBJECTS TREATED BY GROUP DESENSITIZATION ($N = 10$)

Scale	Pre_1		Pre_2		Post	
	M	*SD*	*M*	*SD*	*M*	*SD*
PRCS	21.0	4.38	21.1	2.84	11.5	4.46
SR-Speech	47.2	4.28	49.5	5.78	37.7	5.00
SR-Exam	43.8	9.52	45.5	9.46	38.5	8.89
SR-Interview	36.7	8.43	40.1	8.67	30.8	7.07
SR-Contest	36.1	8.58	36.8	7.18	30.2	6.08
IPAT Anxiety	41.8	10.91	42.7	12.25	35.3	8.79
Emotionality	21.1	4.70	20.4	6.02	17.5	5.45
Extroversion-Introversion	12.4	6.10	15.7	8.38	19.7	8.10

Since the sample was composed of subjects whose major problem consisted of interpersonal performance anxiety, the specific anxiety scales of the battery offer the most meaningful evaluation (Sarason, 1960). Table 2 shows a significant Pre_1-Pre_2-Post main effect for all five specific anxiety scales, indicating movement over the two time periods. Except for SR-interview, which only approached significance, a significant linear component was obtained for all specific anxiety scales. The linear component reflects the lack of change in specific anxiety scores over the untreated wait period, as seen in Pre_1-Pre_2 means presented in Table 4. Of more interest in Table 2 is the significant quadratic component, or curvature, found for every specific performance anxiety scale except SR-contest. This significant deflection is reflected in the reduction of mean anxiety scores over the treatment period (see Table 4). This indicates a significant decrease in reported anxiety following group treatment as compared with changes over the equivalent time period without treatment.

The data presented in Table 3 reveal a significant Pre_1-Pre_2-Post main effect for each of the three general scales as well; however, in contrast to the specific scales, only the linear component was found to be significant. Inspection of means (see Table 4) shows a slight increase in Extroversion, and decrease in Emotionality, to have occurred over the wait period, and continued linearly over the treatment period. The trend in IPAT anxiety scores follows that of the specific anxiety scales, but the quadratic component indicating treatment gain reached only the 10% level of significance.

In general, analysis of self-report scales suggests that treatment by combined group desensitization resulted in: a significant reduction in interpersonal performance anxiety as reported in a number of stressful situations; a degree of general anxiety reduction, and continued increases in gains reflected by increased Extroversion and decreased Emotionality scores. The question of major importance may then be asked—that is, are these reported gains reflected in the objective, extra-experimental criterion of effectiveness, the GPA?

The Academic Performance (GPA) of Untreated Controls and Group Desensitization Ss

To evaluate the effects of the group treatment on external performance, the GPAs for the semester prior to and following the treatment period were obtained for subjects receiving group desensitization and matched controls. A triple-classification analysis of variance (treatment, pre-post GPA, subjects) was performed (see Table 5). If the anxiety reduction reportedly experienced by treated subjects increased their "real-life" effectiveness, it would be anticipated that treated subjects would receive a relatively higher GPA during the second semester than the matched control subjects (Paul & Eriksen, 1964). Such an effect would result in a significant treatment by pre-post interaction.

TABLE 5

ANALYSIS OF VARIANCE ON PRE-POST GRADE POINT AVERAGES FOR SUBJECTS TREATED BY GROUP DESENSITIZATION AND MATCHED CONTROLS ($N = 10$ IN EACH GROUP)

Source	*df*	*MS*	*F*
Treatment	1	.927	—
a. error	18	1.538	
Pre-Post	1	.753	2.87
Treatment × Pre-Post	1	4.686	17.89**
b. error	18	.262	

** $p < .01$.

TABLE 6

MEAN PRE-POST GPAS FOR SUBJECTS TREATED BY GROUP DESENSITIZATION AND MATCHED CONTROLS

Group	Semester GPA			
	1. Pretreatment		2. Posttreatment	
	M	*SD*	*M*	*SD*
Treatment ($N = 10$)	3.152	.5182	3.562	.4248
Control ($N = 10$)	3.532	.8857	2.573	1.4161

As seen in Table 5, the treatment by pre-post interaction for GPA was found to be highly significant, while neither treatment nor pre-post main effects approached significance. The means presented in Table 6 show that the obtained interaction is in the expected direction. The subjects treated by combined group desensitization show a gain in GPA from first to second semesters, whereas the untreated controls show a loss of almost a full grade point.

These results indicate that combined group desensitization did effect significant changes in the extra-experimental performance of treated subjects, adding support to the changes found in the self-report data. It is, therefore, reasonable to compare changes following group desensitization with those obtained over a comparable time period following individual desensitization, insight-oriented psychotherapy, and attention-placebo treatment in the previous study (Paul, 1965).

Comparison of Group Desensitization with Individual Treatments

For purposes of comparison, the mean reduction for each personality and anxiety scale common to all treatment groups is presented in Figure 1. Since subjects treated by group desensitization had no common reference speech for the PRCS, the PRCS and SR-speech scales have been combined into a standardized composite for this comparison (see Paul, 1965). The data for Extroversion-Introversion have been reflected to show decrease in introversion, rather than increased extroversion, for consistency of presentation.

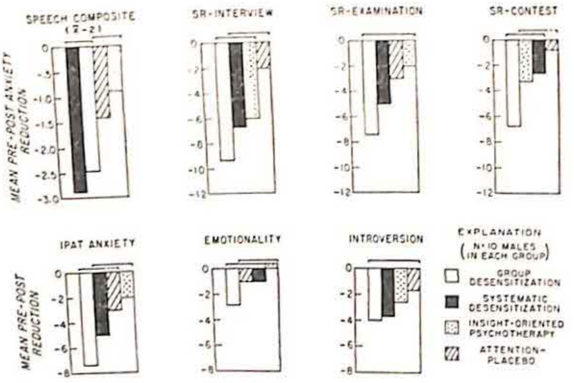


FIG. 1. Mean pre-post reduction on personality and anxiety scales for 4 treatment groups (groups not connected by solid line significantly different at .05 level). (Dotted line indicates $p < .10$.)

Since the data and t tests of differences among the individual treatments, and individual subject data for these groups have been reported in detail in the earlier study, our concern here will be only in the evaluation of group desensitization with the comparable subjects of the three individual treatment groups.

Inspection of Figure 1 reveals significantly greater anxiety reduction on the speech composite for subjects treated by group desensitization than for either attention-placebo or insight-oriented psychotherapy groups. This finding is of central importance since the speaking situation was the prime focus of treatment in all groups. Differences on the speech composite were not significant between the individual and group application of systematic desensitization, even though the group treatment was *not* aided by concurrent enrollment in the speech course.

Tests of significance on the other scales of the test battery found group desensitization to produce significantly greater anxiety reduction than the attention-placebo treatment on the specific interview and contest scales, and differences approaching significance ($p < .10$) on the general anxiety scale. Changes found for the group treatment were also significantly greater than insight-oriented psychotherapy on IPAT anxiety, and Emotionality, even though these scales might be expected to reflect changes typically considered the realm of traditional interview therapy (Frank, 1961).

The one scale on which combined group desensitization would have been predicted to show greatest relative improvement, due to specific focus in treatment, was the SR-Examination scale. As Figure 1 indicates, between-group gains were not significantly different for this scale when evaluated by appropriate t test. Significant differences in variance between the improvement scores of subjects treated by group desensitization and both insight-oriented psychotherapy and attention-placebo groups, combined with small Ns, suggest that the parametric test of differences may be somewhat inappropriate for this scale. Application of a nonparametric test of significance (Kruskal-Wallis One-Way Analysis of Variance on Change Scores) showed the obtained differences to be highly significant ($H = 123.67$, $p < .001$). No significant t tests of differences between the group desensitization and other treatments were found on other scales of the test battery. Inspection of Figure 1 shows that changes obtained for the group treatment equalled or excelled those obtained for individual treatments on every scale. The coefficient of concordance over all scales was found to be highly significant ($W = .74$, $p < .01$), with the relative order of gain following: combined group desensitization, individual desensitization, insight-oriented psychotherapy, and attention-placebo treatment.

Although client self-ratings of improvement have previously failed to discriminate between treatments (Paul, 1965), direct statements of perceived improvement are still of clinical interest. Mean ratings of improvement specific to speaking situations, examinations, and "other areas" obtained on the final test battery for subjects treated by group desensitization were as high or higher than those for any individual treatment, falling at "Much—Very Much Improved" for the speech situation, and "Much Improved" for examination and other areas.

DISCUSSION

The results of the present study demonstrate that the method of systematic desensitization previously found effective in individual treatment (Paul, 1964, 1965) can be efficiently combined with group discussion and administered in groups without loss of effectiveness in the treatment of interpersonal performance anxiety. Analysis of both self-report measures, and the public behavioral criterion of GPA indicate significant reduction in maladaptive anxiety and increased extra-treatment effectiveness. When these results are compared with changes obtained for comparable subjects treated by individual programs of systematic desensitization, insight-oriented psychotherapy, and nonspecific attention-placebo techniques, the combined group desensitization treatment was superior to both the individual insight-oriented and attention-placebo programs. There were no significant differences in the effects of individual desensitization and the group desensitization on individual scales, even though such group desensitization was administered in the *absence* of concurrent enrollment in a speech course.

The original work with group desensitization (Lazarus, 1961) was concerned with the efficacy of group desensitization versus the traditional method of interpretation of "group dynamics" in the treatment of specific phobic disorders. Because of differences in subjects and design, a direct comparison between the present results and those of any one of Lazarus' four groups is neither intended nor justified. The present results do suggest, however, that group desensitization need not be limited to treatment of specific phobias, but that it is also effective in treating more generalized social-evaluative anxiety.

While a control group receiving only traditional group methods was not included in the present investigation, previous group studies have failed to obtain significant improvement on extra-treatment indices of functioning such as GPA (Baymur & Patterson, 1960; Broedel, Ohlsen, Proff, & Southard, 1960; Fiedler, 1949; Spielberger & Weitz, 1962). The effectiveness of group desensitization is enhanced upon comparison of the present results with those obtained by Spielberger and Weitz. The subgroup of nine of their subjects who did show an increase in GPA following treatment by traditional group methods was described as "active-repressive," on the basis of MMPI profiles elevated on *Ma* and *Hy* scales, in addition to high Taylor MAS scores, that is, anxious-extroverts. Spielberger and Weitz also identified a subgroup of subjects who were described as "passive, withdrawn, and immature" on the basis of elevated *Mf*, *Sc*, and *Pt* on the MMPI, as well as high MAS. This subgroup of anxious-introverts appears most like the subjects included in the present study. In contrast to their anxious-extrovert subgroup, these subjects actually showed a slightly *less* improvement on GPA than untreated controls. Thus, while a controlled *comparative* investigation of these differing group methods is desirable, results to date suggest that combined group desensitization may be a more effective group treatment of social-evaluative anxiety.

A detailed discussion of possible factors operating within various approaches to short-term treatment is presented in Paul's (1965) monograph. Although the effects of combined group desensitization were not significantly different from individual desensitization on any specific scale, the consistency of anxiety reduction found for the group method over all scales suggests the operation of additional factors. These factors seem to reflect the effects of both group discussion, and the experience of systematic desensitization in a group rather than individual setting. The examples of "other areas" in which subjects treated by group desensitization reported improvement suggest a combination of the specific examples previously found for systematic desensitization alone, and the more abstract examples typical on insight-oriented psychotherapy (Paul, 1965)—for example, such specifics as: "face to face interaction," "talking to friends and relatives," "talking in social groups," and "informal conversation," to more generalized examples as: "personal problems," "general self-confidence," "general composure," and "general tension."

Additionally, both therapists noted that subjects in the group seemed to perceive the desensitization method as an active mastery technique which they could acquire and use themselves, more than in the individual application. Client's descriptions of utilizing desensitization training to master anticipated areas of stress themselves suggest the development of a confidence-building "how to cope" orientation.

One factor which may add additional benefit over individual desensitization derives from the construction of hierarchies through group discussion. Not only may discrimination learning occur as in individual programs, but a reduction in anxiety as subjects become aware of the similarity between themselves and other group members. Lazarus (1961) and Murray (1964), among others, have speculated that anxiety undergoes some measure of extinction in the process of such group interaction, and that interpersonal situations may evoke responses similar to those of deep muscle relaxation. Lazarus found fewer sessions were required to desensitize his phobic subjects after unsuccessful attendance in his other groups, suggesting an addictive effect. Similarly, one to two sessions were sufficient to complete a test anxiety hierarchy in the group, after earlier group experience in the present study, while similar hierarchies have required three to five individual sessions without prior group experience (Paul, 1964).

The group setting also provides an ideal opportunity for persons debilitated by social anxiety to practice deficient skills after a degree of anxiety reduction has been attained. Initially, the group treatment appeared to increase immediate feelings of awkwardness. However, after a few sessions the groups pulled together, as described in the accumulated literature on group psychotherapy (Rosenbaum & Berger, 1963), typically left the clinic together, and reported socializing between sessions. Within treatment, the lively discussion of the training and its effects provides immediate positive reinforcement to changes in behavior and attitudes not possible in individual treatment.

The failure to find significant differences attributable to the two therapists lends support to the evidence that the methods themselves primarily influence outcome in this form of treatment. However, the therapists were very similar in attitudes and techniques, and both therapists were experienced with the modified desensitization method. Nevertheless, promise of an effective method of treatment which can be taught and practiced in a systematic, reliable manner is reassuring in the face of the demand for therapeutic services, (Geer, 1964) and the current state of affairs in the area of psychotherapy (Clark, 1965). The continued reports of success with desensitization procedures suggest not only the obvious research emphasis needed to delineate the parameters of influence (e.g., Lang & Lazovik, 1963), but call for greater interest in outcome research to determine the extent and limits of applicability and long-term assessment of duration of effects.

As with all studies, there are limitations to the present findings, and to the extent to which results can be generalized. The number of subjects was small. However, the use of careful matching procedures enhances the meaningfulness of the data. While a comparative outcome study of group-therapy methods is needed, the purpose of the present study was *not* to pit group desensitization against traditional methods. Rather, its purpose was to evaluate the feasibility and effectiveness of administering modified systematic desensitization in a group. It should also be noted that the major evaluation of the group method involved the own-control wait period, and the matched untreated controls. The individual treatment data were included to provide a basis for evaluating the extent of effects of the group treatment, but are not directly comparable since individual treatments consisted of 5 hours of therapy contact, while group subjects were provided a total of 9 hours of therapy contact. However, since the combined group desensitization procedure required an average of less than *two hours* therapist time per client, this short-term, time-limited approach appears to be not only effective, but efficient as well.

REFERENCES

BANDURA, A. Psychotherapy as a learning process. *Psychological Bulletin*, 1961, **58**, 143–159.

BAYMUR, F. B., & PATTERSON, C. H. A comparison of three methods of assisting underachieving high school students. *Journal of Counseling Psychology*, 1960, **7**, 83–89.

BENDIG, A. W. Pittsburg scale of social extroversion-introversion and emotionality. *Journal of Psychology*, 1962, **53**, 199–210.

BROEDEL, J., OHLSEN, M., PROFF, F., & SOUTHARD, C. The effects of group counseling on gifted underachieving adolescents. *Journal of Counseling Psychology*, 1960, **7**, 163–170.

CATTELL, R. B. *The IPAT Anxiety Scale.* Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

CLARK, K. E. (Chairman) Committee on the scientific and professional aims of psychology: Preliminary report. *American Psychologist,* 1965, **20**, 93–100.

ENDLER, N. S., HUNT, J. McV., & ROSENSTEIN, A. J. An S-R inventory of anxiousness. *Psychological Monographs,* 1962, **76** (17, Whole No. 536).

FIEDLER, F. E. An experimental approach to preventative psychotherapy. *Journal of Abnormal and Social Psychology,* 1949, **44**, 386–393.

FRANK, J. D. Problems of controls in psychotherapy. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy,* Washington, D. C.: APA, 1959.

FRANK, J. D. *Persuasion & healing,* Baltimore: Johns Hopkins, 1961.

GEER, J. H. Phobia treated by reciprocal inhibition. *Journal of Abnormal and Social Psychology,* 1964, **69**, 642–645.

HUNT, J. McV. Concerning the impact of group psychotherapy on psychology. *International Journal of Group Psychotherapy,* 1964, **1**, 3–31.

HUSEK, T. R., & ALEXANDER, S. The effectiveness of the Anxiety Differential in examination stress situations. *Educational and Psychological Measurement,* 1963, **23**, 309–318.

LANG, P. J., & LAZOVIK, A. D. Experimental desensitization of a phobia. *Journal of Abnormal and Social Psychology,* 1963, **66**, 519–525.

LAZARUS, A. A. Group therapy of phobic disorders by systematic desensitization. *Journal of Abnormal and Social Psychology,* 1961, **63**, 504–510.

MUENCH, G. The comparative effectiveness of long-term, short-term, and interrupted psychotherapy. Paper read at the Western Psychological Association, Portland, Oregon, April 1964.

MURRAY, E. J. Sociotropic-learning approach to psychotherapy. In P. Worschel & D. Byrne (Eds.), *Personality change.* New York: Wiley, 1964.

PAUL, G. L. Modifications of systematic desensitization based on case study. Paper read at Western Psychological Association, Portland, Oregon, April 1964.

PAUL, G. L. *Insight versus desensitization in psychotherapy: An experiment in anxiety reduction.* Stanford: Stanford Univer. Press, 1965.

PAUL, G. L., & ERIKSEN, C. W. Effects of test anxiety on "real life" examinations. *Journal of Personality,* 1964, **32**, 480–494.

ROSENBAUM, M., & BERGER, M. *Group psychotherapy and group function.* New York: Basic Books, 1963.

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychological Bulletin,* 1960, **57**, 403–415.

SCHOFIELD, W. *Psychotherapy: The purchase of friendship.* Englewood Cliffs, N. J.: Prentice-Hall, 1964.

SHLIEN, J. M., MOSAK, H. H., & DREIKURS, P. Effect of time limits: A comparison of client centered and Adlerian psychotherapy. *University of Chicago Library Counseling Center Discussion Papers,* 1960, **6**, No. 8.

SPIELBERGER, C. D. The effects of manifest anxiety on the academic achievement of college students. *Mental Hygiene,* 1962, **46**, 420–426.

SPIELBERGER, C. D., & KATZMEYER, W. G. Manifest anxiety, intelligence, and college grades. *Journal of Consulting Psychology,* 1959, **23**, 278.

SPIELBERGER, C. D., & WEITZ, H. Group counseling and the academic performance of anxious college freshmen. *Journal of Counseling Psychology,* 1962, **9**, 195–204.

WOLPE, J. The systematic desensitization treatment of neuroses. *Journal of Nervous and Mental Diseases,* 1961, **132**, 189–203.

(Received March 29, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 136–141

AN EXPERIMENTAL STUDY OF THE DOUBLE BIND HYPOTHESIS[1]

EUGENE L. RINGUETTE AND TRUDY KENNEDY[2]

UCLA Neuropsychiatric Institute

The primary aim of the study was to assess whether persons highly conversant with the double bind concept could reliably identify such communication in letters, and, if so, whether its presence or absence was related to the fact of the letter having been written by the parent of a schizophrenic or nonschizophrenic person. The interjudge correlations of the double bind judges were very low, indicating that their judgments regarding presence or absence of double bind communication were not reliable. Other judges achieved statistically significant interjudge reliability and were able to differentiate between letters written by parents of patients and a comparison group. It was concluded from these results that a more critical appraisal of the double bind concept is indicated.

The double bind hypothesis states, in brief, that schizophrenic symptoms are reactions to an inability to judge exactly the meaning of communications which are received from others. It is postulated that this difficulty results from an inability to assess properly the context within which information is received, and, hence, the meaning or interpretation of the information is uncertain. It is further postulated that this inability to understand and discriminate among the abstract aspects of communication is caused by prolonged exposure to an environment in which simultaneous communications at various levels of abstraction are consistently incongruent. Such communication, in which incongruent messages are received at different abstract levels, is termed "double bind" communication (Bateson, Jackson, Haley, & Weakland, 1956).

It is also postulated that double bind communication is a consistent and integral aspect of communication within the families of schizophrenics, as well as in letters from mothers to schizophrenic offspring, and is regarded as a factor in the etiology of the disorder.

> Our hope is that these operations will provide a clearly evident record of the continuing, repetitive double binding which we hypothesize goes on steadily from infantile beginnings in the family situation of individuals who become schizophrenic. This basic family situation, and the overtly communicational characteristics of schizophrenia, have been the major focus of this paper [Bateson et al., 1956, p. 262].

> Finally, we may note again that both the major pattern of communication discerned in these letters and its effects on their recipients agree closely with our research group's original and prior concept of the double bind as a communicational pattern and our hypothesis of its relationship to schizophrenia and its etiology [Weakland & Fry, 1962, p. 623]. It is found that: 1) While the letters vary greatly in details of content, style, etc., they exhibit similar pervasive and highly influential patterns of incongruent communication. 2) These letters agree with another schizophrenic's characterization of such letters generally. 3) The observed pattern fits prior general statements of the authors' research group about the "double bind" and incongruent communication in schizophrenia [Weakland & Fry, 1962, p. 604].

It is further stated that letters written by the families of schizophrenics are especially suited to analysis of double bind communication in that they are objective, free of extraneous vocal, facial, and gestural messages,

---

[1] The study was aided by a grant from the Medical Staff Association Fund, Department of Psychiatry, Center for the Health Sciences. Appreciation is also expressed to O. L. Gericke, Superintendent, and the staff of Patton State Hospital for their help in securing material for the study.

Special acknowledgment and gratitude is extended to the following persons who participated as judges in the study: Gregory Bateson, James H. Bryan, Alexander B. Caldwell, William F. Fry, Ralph Greenson, Peter B. Gruenberg, John Hanley, Edward J. Kollar, Ronald Sager, Lowell H. Storms, Charles W. Tidd, and Paul Watzlawick.

[2] Now a Research Fellow, University of Melbourne.

and are condensed and unitary examples of the writer's communicative habits.

> It is therefore reasonable to assume that while a letter, in comparison with ordinary speech, is a form easier for analytic study, it remains typical of the sender's communicative habits in major respects, and indeed may highlight these because of the selectivity which the writer has exercised [Weakland & Fry, 1962, p. 605].

A review of the double bind literature, then, indicates that, as a theory, the double bind hypothesis postulates the following: (*a*) The "double bind" is a definable phenomenon of communication. (*b*) It involves a major communicational pattern within the families of schizophrenics and is a factor in the etiology of the disorder. (*c*) The letters of mothers of schizophrenics to their schizophrenic offspring (although the phenomenon may not be restricted to mothers' communication) exhibit discernible pervasive patterns of incongruent communication, that is, evidence of double bind communication. (*d*) Such letters are particularly suited to analysis of double bind communication.

In view of the above, the purpose of the present study was to assess whether persons highly conversant with the concept of the double bind, in the sense of being significantly associated with its formulation or elaboration, could agree with regard to its presence or absence in communication. Its presence in letters written by parents of schizophrenics is, of course, assumed by the double bind hypothesis.

A second consideration relating to the reliability of the construct was whether or not it could be reliably identified by persons trained to do so, because if this were not possible the applicability of the double bind hypothesis would be restricted to only a few expert persons.

In addition to the above, there were certain secondary considerations which were contingent upon the reliability with which double bind communication could be identified. If the double bind were a reliable phenomenon in letters, the expectation derived from the double bind hypothesis would be that it should occur more frequently in letters written by parents of schizophrenics than in other letters. In an effort to evaluate this eventuality, two comparison groups of letters, one written by parents of nonschizophrenic psychiatric patients and one written by hospital volunteers, were included. Finally, other types of judges, judging the letters on variables other than presence-absence of double bind communication, were included in order to evaluate the power of the double bind hypothesis to discriminate between schizophrenia and nonschizophrenia, relative to other measures such as clinical judgment, which might be applied to the letters. A consideration of the latter, again, depends upon a demonstration that the judged presence-absence of double bind communication is a reliable phenomenon in the letter sample, so the primary aim of the study was specifically to assess the reliability with which double bind communication in letters could be identified.

METHOD

Sixty letters were used; 40 were letters which had been received by patients in a state hospital from their parents, and 20 were letters written by hospital volunteers as if they were writing to offspring hospitalized at the state hospital from which the patient letters were obtained. Half of the patients whose letters were used had a hospital diagnosis of schizophrenia, and half had nonschizophrenic diagnoses. The criteria used for including a patient's letter in the sample were that the patient be between the ages of 18 and 30 years, that this be his first known admission, that he had been hospitalized for less than 6 months, and that he be Caucasian and English-speaking. The volunteers who wrote letters ranged in age from 38 to 62, and had no children who had been hospitalized for psychiatric reasons. They were used as a comparison group because of the fact that letters received by nonpsychiatric patients were found to be easily differentiated from psychiatric patient letters on the basis of content. This group of letters written by volunteers, then, was not presumed to constitute a control group; rather, it is a second comparison group, in addition to the letters received by nonschizophrenic patients, in which double bind communication would not be expected to exist according to the terms of the double bind hypothesis.

In summary, then, there were 60 letters used in the study, of which 20 had been received by schizophrenic patients from their parents, 20 had been received by nonschizophrenic patients from their parents, and 20 had been written by volunteers as if they were writing to hospitalized offspring.

All letters were presented in mimeographed typewritten form to the judges participating in the study, and the letters were presented in a different randomized order to each judge.

TABLE 1

DISTRIBUTIONS OF THE RATINGS OF THE EXPERT, TRAINED, AND NAIVE JUDGES

Judges	Rating scale						
	1	2	3	4	5	6	7
Expert 1	5	16	2	12	7	7	11
Expert 2	4	6	1	8	16	18	7
Expert 3	7	11	4	5	4	7	22
Trained 1	18	12	2	5	8	5	10
Trained 2	15	10	4	0	6	8	17
Trained 3	0	15	1	7	6	17	14
Naive 1	3	16	25	5	3	6	2
Naive 2	4	14	14	13	7	7	1
Naive 3	1	9	15	11	6	14	4

Five types of judges were used in the study, with three judges in each type. The "expert double bind" judges were persons closely involved in the formulation of the double bind hypothesis and, hence, highly conversant with the concept. These judges were told that "the attached letters were written *by parents* to their adult or adolescent *offspring*. Some of these offspring are schizophrenic and some are not, but the numbers of schizophrenic and nonschizophrenic offspring are not equal." They were then asked to rate each letter on a 7-point scale with regard to the extent to which it suggested the presence or absence of double bind communication, with one representing presence and seven representing absence. They were not asked to judge whether each letter had been received by a schizophrenic or nonschizophrenic person.

A second group of judges was termed the "trained double bind" type. These were first-year psychiatric residents whose familiarity with the double bind concept was gained in a formal seminar and in a subsequent meeting with the authors of the present study. These judges read selected articles relating to the concept and then were asked to make the same judgments as were made by the expert judges except that they were instructed to use the specific criteria for identifying double bind communication which had been enumerated by Weakland and Fry (1962).

A third category of judges was termed the "uninformed clinicians." These were experienced clinicians who were given the same information regarding the letters as had been given the expert and trained judges. They were asked to judge whether each letter had been written by the parent of a schizophrenic or nonschizophrenic person.

A fourth type of judges was termed the "informed clinicians." This group was comprised of experienced clinicians who were informed regarding the origin of the letters and were instructed to sort the letters into their respective groups, that is, those received by schizophrenic patients, those received by nonschizophrenic patients, and those written by volunteers.

The fifth group of judges was termed the "naive" type. These were persons of comparable educational level to the other judges but not associated with medicine or the social sciences. They were instructed to rate each letter on a 7-point scale, with one representing "like" and seven representing "dislike," with regard to "whether your general reaction to the letter is a positive or negative one; that is, is it a letter you would or would not like to receive."

RESULTS

The frequencies with which each judge in the expert, trained, and naive judge types used each point on their respective 7-point scales are shown in Table 1.

The average interjudge correlations for these judge types were: expert = .19, trained = .26 ($p < .05$), and naive = .39 ($p < .01$). In view of the low interjudge correlations of the expert double bind judges, interjudge correlations were calculated separately for their ratings of the 20 letters written by mothers of schizophrenics on the chance that they had achieved some agreement on those letters which was obscured by lack of agreement on the two comparison groups of letters. The three interjudge correlations for the expert judges' ratings of the letters written by mothers of schizophrenics were .14, .27, and .17, none of which is significant, and which again average .19.

Interjudge contingency coefficients and chi-square tests of significance were computed for the uninformed and informed clinician judges. The average coefficient for the uninformed clinicians was .13, while that for the informed clinicians was .44. The chi-squares obtained for the informed clinicians were all significant well beyond the .05 level of confidence, and the coefficient for the informed judges seems particularly high in view of the fact that the upper limit of a contingency coefficient for a 3×3 table is .82 (Siegel, 1956).

The results indicate, then, that the judgments of the informed clinicians, naive judges, and, to a lesser extent, the trained double bind judges reflect statistically significant interjudge reliability, while those of the expert double bind judges and uninformed clinicians do not.

An analysis of variance was performed with diagnostic groups (D) and judge types (T) as the two factors. The three levels of D were schizophrenic letters, nonschizophrenic letters, and volunteer letters, while the three levels of T were expert, trained, and naive judge types.

A significant effect was obtained due to letters within diagnostic groups which reflects the fact that mean ratings, over all judges, of the letters differed within each diagnostic group. In effect, this suggests that the ratings were not randomly assigned. A significant effect due to judges within judge types also was obtained which indicated that the judges within each judge type differed with respect to the mean ratings which they assigned; that is, within each judge type, the judges differed in terms of the part of the rating scale which they tended to use. This trend may also be seen in the distribution of judges' ratings presented above (see Table 2).

The diagnostic groups did not differ significantly with regard to mean ratings received, nor did the judge types differ significantly with regard to mean ratings assigned. There was, however, a significant interaction effect between diagnostic groups and judge types which indicated that there were significant differences between the mean ratings assigned to the diagnostic groups as a function of judge type. The mean ratings of diagnostic groups by judge types are given in Table 3.

Table 3 illustrates the failure of any of the judge types to differentiate between the schizophrenic and nonschizophrenic letter groups. The only significant difference found was for the naive judges who assigned a lower mean rating to the volunteer letters than to the schizophrenic or nonschizophrenic letters ($t = 2.68$, $df = 19$, $p < .02$).

TABLE 2

ANALYSIS OF VARIANCE SUMMARY OF THE RATINGS OF THE EXPERT DOUBLE BIND, TRAINED DOUBLE BIND, AND NAIVE JUDGES

Source	*df*	*MS*	Comparison		*F*
			df	*MS*	
D	2	12.20	57	7.15	1.71
T	2	41.11	7.26[a]	20.00	2.05
L(D)	57	7.15	342	2.84	2.52**
J(T)	6	18.15	342	2.84	6.39**
DT	4	11.78	114	4.96	2.51*

[a] Equivalent degrees of freedom for the judge types comparison mean square calculated according to Bennett and Franklin (1954), p. 368.

* $p < .05$.

** $p < .01$.

TABLE 3

MEAN RATINGS ASSIGNED TO THE VOLUNTEER, NONSCHIZOPHRENIC, AND SCHIZOPHRENIC LETTERS BY THE EXPERT DOUBLE BIND, TRAINED DOUBLE BIND, AND NAIVE JUDGE TYPES

Letters	Type			
	Expert	Trained	Naive	
Diagnosis				
Volunteer	4.78	3.83	2.83*	3.82
Nonschizophrenic	4.37	4.27	4.12	4.25
Schizophrenic	4.62	4.28	3.95	4.28
Total	4.59	4.13	3.63	4.11

The judgments of the uninformed and informed clinicians, who were treated separately, were cross-tabulated with the diagnostic groups to which the letters belonged. These cross-tabulations, pooled for illustrative purposes, are presented in Tables 4 and 5.

TABLE 4

CROSS-TABULATION OF THE POOLED JUDGMENTS OF THE UNINFORMED CLINICIANS AND THE DIAGNOSTIC GROUPS OF THE LETTERS

Letters	Judgments		
	Nonschizophrenic	Schizophrenic	
Diagnosis			
Volunteer	38	22	60
Nonschizophrenic	9	51	60
Schizophrenic	12	48	60
Total	59	121	180

TABLE 5

CROSS-TABULATION OF THE POOLED JUDGMENTS OF THE INFORMED CLINICIANS AND THE DIAGNOSTIC GROUPS OF THE LETTERS

Letter	Judgments			
	Volunteer	Nonschizophrenic	Schizophrenic	
Diagnosis				
Volunteer	34	18	8	60
Nonschizophrenic	12	20	28	60
Schizophrenic	9	24	27	60
Total	55	62	63	180

Inspection of these cross-tabulations clearly illustrates that the clinician judges were unable to differentiate between the schizophrenic and nonschizophrenic letters. It is equally clear, however, that the clinicians' judgments with regard to the volunteer letters appear to be considerably more valid. Two-by-two tables comparing each uninformed clinician's judgment of schizophrenic or nonschizophrenic with whether the letters belonged to the volunteer or patient groups were formed. Similar tabulations were done for the informed clinicians in which their judgments of volunteer and patient were compared with whether the letters belonged to the volunteer or patient groups. The chi-square test was performed for each clinician judge, and each obtained chi-square value was significant at the .01 level of confidence.

In summary, it appears that none of the judge types was able to differentiate between the letters received by schizophrenic patients and those received by nonschizophrenic patients. With the exception of the expert and trained double bind judges, however, the ratings of all of the other judges significantly differentiated the volunteer from the patient letters.

Because of the fact that none of the judge types successfully differentiated between the letters received by schizophrenics and those received by nonschizophrenics, the authors were prompted to pursue some a posteriori investigation. Using the Thorndike-Lorge word frequency dictionary (1944), the percentage of infrequent verbs was calculated for each letter. The median percentages of infrequent verbs for the volunteer, schizophrenic, and nonschizophrenic letters were 7.5%, 6.5%, and 4.0%, respectively. Application of the Extension of the Median Test (Siegel, 1956) produced a χ^2 value of 6.61 ($p < .05$).

The mean number of words of the letters in each group was also calculated and found to be 140.95 for the volunteer letters, 151.20 for the nonschizophrenic letters, and 195.30 for the schizophrenic letters. Although the greater mean length of the schizophrenic letters is notable, the F ratio obtained from a one-way analysis of variance was 2.21, which falls short of the value of 3.15 representing significance at the .05 level of confidence. Thus, length in itself does not appear to constitute a cue for differentiating the letter groups.

DISCUSSION

According to the terms of the double bind hypothesis, the "double bind" is a definable and identifiable phenomenon in communication and has a causal, though complex, relationship to schizophrenia. In addition, the theory postulates that it is a consistent aspect of communication within schizophrenic families and is present and discernible in letters written by mothers of schizophrenics. To paraphrase Weakland and Fry's (1962) statement regarding "specific connections between data and theory" [p. 623], letters are a ground upon which data and the double bind hypothesis meet.

Accepting the propositions of the theory, the aim of the present study was to ascertain whether persons truly knowledgeable with regard to the double bind formulation could reliably identify this phenomenon in a form of communication which had been proposed as being especially suited for this by proponents of the double bind hypothesis. The results clearly indicate that they were not able to do this. It is axiomatic, of course, that the validity of a phenomenon can be neither verified nor refuted if reliable measurement is lacking. In short, unless a phenomenon can be reliably identified, it cannot be considered to be a meaningful or valid phenomenon.

The failure of the expert double bind judges to achieve reliability in the present study seems especially compelling in view of the fact that they represented a select and relatively large sample. As "expert" was defined in the study, a perusal of the literature on the double bind hypothesis indicates that perhaps six to eight persons could claim such expertness. Thus, the sample of expert judges constituted one-third to one-half of the population of such persons.

In view of the present results obtained using a sample of letters, one might propose the use of films, tape recordings, or a series of such communications as alternative media especially suited to such analysis. Such a proposal, however, is a question for empirical verification rather than debate.

In spite of the negative results with regard to the double bind, other variables showed promise of reliability. The task of the informed clinician judges was such that significant reliability was not surprising. Their task was essentially less difficult and more restricted. Of more interest, perhaps, is the fact that the like-dislike dimension was a meaningful one.

The results of the study prompt the general conclusion that the double bind hypothesis requires more critical appraisal than it has had. It is a theoretical formulation which has not been anchored in data, and, taken on its stated ground in the present instance, the double bind appears to be an unreliable phenomenon or a nonexistent one. The theory postulates that double bind communication is present and discernible in letters; yet, persons most knowledgeable regarding the "double bind" cannot reliably identify it in letters. The possible reasons for this are: (*a*) double bind communication is not present in letters, in which case a postulate of the theory is clearly invalid; (*b*) it is not presently a measurable phenomenon; (*c*) it actually does not exist.

REFERENCES

BATESON, G., JACKSON, D., HALEY, J., & WEAKLAND, J. Towards a theory of schizophrenia. *Behavioral Science,* 1956, 1, 251–264.

BENNETT, C., & FRANKLIN, N. *Statistical analysis in chemistry & the chemical industry.* London: Wiley, 1954.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teacher's College, Columbia University, Bureau of Publications, 1944.

WEAKLAND, J., & FRY, W. Letters of mothers of schizophrenics. *American Journal of Orthopsychiatry,* 1962, 32(4).

(Received April 2, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 142–146

THE RELATIONSHIP BETWEEN A MEASURE OF TRANSITORY ANXIETY AND SPONTANEOUS AUTONOMIC ACTIVITY[1]

EDWARD S. KATKIN

State University of New York at Buffalo

The effects of stress (threat of shock) on GSR nonspecific responses were evaluated for Ss who differed in transitory anxiety as measured by the Affect Adjective Check List (AACL). 52 Ss were equally divided among 4 groups: Hi AACL Stress, Lo AACL Stress, Hi AACL Nonstress, and Lo AACL Nonstress. First, all Ss rested; then both Stress groups were threatened with shock, while both Nonstress groups continued resting; finally, all Ss rested again. During the 2nd period GSR nonspecifics increased for both stress groups; however, during the final period, after the cessation of stress, GSR nonspecifics decreased more for the Lo AACL Ss than for the Hi AACL Ss. It was concluded that the AACL may be related to "autonomic recovery rate."

Anxiety is currently regarded as a central explanatory construct by many personality theorists, and it is often assumed that an appropriate conceptualization of it is essential to a full understanding of personality. In order to use anxiety as an explanatory construct, however, objective and reliable measures of the properties attributed to it must be developed. Among various approaches to the assessment of anxiety, autonomic nervous system indices and self-report inventories have been frequently employed (see Martin, 1961).

One measure of autonomic activity which has been employed as a physiological correlate of anxiety is GSR nonspecifics, an index of fluctuations in skin resistance which occur in the absence of specific stimulation (Cohen, Silverman, & Burch, 1956; Greiner & Burch, 1955). This measure of spontaneous autonomic activity has been shown to be an effective index of emotional response to stress (Katkin, 1963, 1965), as well as a valid measure of chemically induced "activation" (Greiner & Burch, 1955).

Among self-report indices of anxiety, none has received more attention than the Taylor Manifest Anxiety (*MA*) Scale (Taylor, 1953), which has been demonstrated to be related to clinical judgments of anxiety (Eriksen & Davids, 1955; Hoyt & Magoon, 1954) as well as to other scalar indices of anxiety and general neuroticism (Bendig, 1957; Cattell & Scheier, 1958). However, attempts to relate the *MA* scale to physiological indices such as the GSR have produced largely negative results (Katkin, 1965; Sarason, 1960; Silverman, 1957). One possible reason for the repeated failure to find relationships between *MA* scale scores and GSR responses to laboratory situations may be that the *MA* scale items refer frequently to general states of the organism (e.g., *I work under a great deal of tension*), rather than emotional reactions to transitory situations. Recently, Zuckerman (1960) has described an Affect Adjective Check List (AACL), which allows measurement of anxiety at a given moment. Furthermore, it has been demonstrated that increases in AACL scores are associated with transitory fears of failure on classroom examinations (Zuckerman & Biase, 1962) and with emotional responses to a film describing prefrontal lobotomy (Winter, Ferreira, & Ransom, 1963).

The goal of the present study was to evaluate the relationships between scores on the AACL and GSR nonspecifics during differentially stressful conditions. It was expected that under stressful conditions all subjects would show greater autonomic arousal than comparable subjects who were not stressed, but that subjects in the Stress condition with high AACL scores would show greater autonomic arousal than those with low AACL scores.

---

[1] This investigation was supported in part by a Public Health Service Fellowship (MF-17,490) from the National Institute of Mental Health. The author is indebted to Jerry F. Cataldo for his assistance in analyzing data.

Portions of this paper were presented at the meetings of the Midwestern Psychological Association, 1965.

METHOD

Subjects

Subjects were 62 male undergraduates at Duke University. All subjects were volunteers and received credit towards a course requirement for their participation. In the course of running the experiment 10 *Ss* were eliminated because of apparatus failure and procedural difficulties.

Apparatus

The experimental chamber was a well-lighted, sound-resistant room with constant temperature. It contained a comfortable reclining chair for subjects, a chair for the experimenter, a tape recorder, and various equipment necessary for the attachment of GSR electrodes. The apparatus for amplifying and recording GSR responses was located outside the chamber. A detailed description of the GSR detection and amplification system has been presented elsewhere (Katkin, 1965).

Procedure

Subjects were randomly assigned to stress and nonstress conditions consisting of three 10-minute time periods. The initial and final periods were structured as rest periods for all subjects; however, in the middle period, subjects in the stress condition were threatened with electric shock, while subjects in the nonstress condition continued to rest. Just before the initial rest period began, the subjects were asked to fill out the AACL. These were not scored until the completion of the experiment, thereby preventing the experimenter from knowing the AACL scores of the subjects while he conducted the experiment.

First experimental period. At the beginning of the experiment all subjects were given the following instructions:

> The object of this study is to examine certain physiological reactions while you relax. In order to do that I am going to ask you to just sit back and relax for a while, as I measure some of your physiological reactions. The measure I am specifically interested in is your skin resistance. (At this point the subject was shown the GSR electrodes, and they were attached to his finger and arm.) These measure the resistance of your skin. Do you have any questions? (The experimenter paused to answer any questions which may have arisen.) Now before we actually get started with the recording, we would like you to fill out this check list. It is very simple. Just check those words which describe how you feel *today*. Some of the words may sound alike, but we want you to check all the words that describe your feelings *today*. (At this point the subjects were given the AACL to fill out.) Do you have any questions? (The experimenter paused to answer any questions which may have arisen.) All right, then, just relax. (At this point, the experimenter tilted the subject's chair into its reclining position.) I will come back in about 10 minutes to give you further instructions and see if you are comfortable. After 10 minutes had elapsed, the experimenter returned to the chamber.

Second experimental period. The instructions given the subjects in the stress group at the beginning of this period were designed to convince them that they were to receive electric shocks, and that they would have no control over the delivery of the shocks. These instructions have been described in detail in an earlier report (Katkin, 1965). No shocks were actually delivered at any time to any of the subjects who participated in the experiment.

The instructions given subjects in the nonstress condition at the beginning of the second period were designed to convince them that the experimenter was interested only in physiological responses during rest. The time spent between periods and the amount of physical activity of subjects in the stress and nonstress conditions were equated in order to control for changes in GSR activity resulting from physical activity or time.

Third experimental period. At the beginning of this period the subjects in the stress condition were told: "This period will be just like the initial period." The subjects in the nonstress condition were told: "This period will be another 10-minute period of relaxation." At the conclusion of this final rest period the subjects were thanked for their cooperation, and, in a conversational manner, informed that there were actually three experimental groups: a group which received electric shock, a group which was threatened but did not receive electric shock, and a group which was not threatened and did not receive shock. This information was given in order to minimize the possibility that future subjects would be informed that there was no danger of being shocked. Each subject was asked to maintain secrecy about the purpose and design of the experiment so that future subjects would not develop preconceptions about its nature.

RESULTS

A GSR nonspecific was defined as a decrease in skin resistance of at least 100 ohms, followed by a gradual return to predeflection level. A continuous record of GSR nonspecifics was obtained for each subject, and subjects' scores on this measure were determined by counting the number of nonspecifics in each 10-minute time period. In this manner, three GSR nonspecifics scores were obtained for each subject.

TABLE 1

MEANS AND *SD*s OF GSR NONSPECIFICS FOR THE FOUR EXPERIMENTAL GROUPS, FOR EACH 10-MINUTE PERIOD

Group	Rest		Stress		Rest	
	First period		Second period		Third period	
	M	*SD*	*M*	*SD*	*M*	*SD*
Hi AACL Stress	24.54	23.42	60.85	28.23	37.62	23.54
Lo AACL Stress	32.69	22.28	69.15	31.03	31.31	21.65
Hi AACL Nonstress	28.77	20.48	31.77	21.75	35.39	22.74
Lo AACL Nonstress	32.46	27.02	39.92	34.64	34.23	29.28

The subjects were divided into high anxiety (Hi AACL) and low anxiety (Lo AACL) groups on the basis of whether their scores on the AACL were above or below the median for the total sample. Students who scored 8 or above were designated Hi AACL, and those who scored 7 or below were designated Lo AACL. Thus, four groups, each with 13 subjects, were constituted: Hi AACL Stress, Lo AACL Stress, Hi AACL Nonstress, and Lo AACL Nonstress.

The means and *SD*s of the nonspecifics for the four groups for each 10-minute period are presented in Table 1. During the initial rest period all four groups gave approximately the same number of nonspecifics; the tendency for the Lo AACL groups to give a slightly larger number of responses than the Hi AACL groups during the initial rest period was not significant ($t = 0.88$; $df = 50$). During the second experimental period, as may be noted in Table 1, mean GSR nonspecifics for subjects in the Hi AACL Stress and Lo AACL Stress groups increased markedly, while those of subjects in the Hi AACL Nonstress and Lo AACL Nonstress groups increased only slightly. During the final rest period mean GSR nonspecifics of the subjects in the Hi AACL Stress group decreased somewhat, but remained considerably higher than the comparable mean for the initial rest period; mean GSR nonspecifics for the subjects in the Lo AACL Stress group decreased markedly, returning to a point just below that of the initial rest period. Also during the final rest period, mean GSR nonspecifics of Hi AACL subjects in the Nonstress group increased slightly, while those of Lo AACL subjects in the Nonstress group decreased slightly.

Analysis of variance of these data indicated that a significant periods by stress interaction ($F = 24.57$; $df = 2/96$; $p < .001$) and a significant periods by anxiety interaction ($F = 3.22$; $df = 2/96$; $p < .05$) as well as a significant main effect of periods ($F = 37.72$; $df = 2/96$; $p < .001$) were obtained. The significant periods by stress interaction reflects the fact that marked differences in GSR nonspecifics were obtained for subjects in the Stress and Nonstress conditions during the second period, but not during the initial and final rest periods. Similarly, the main effect of periods resulted largely from the obtained differences between subjects in the Stress and Nonstress conditions during the second experimental period. The significant periods by anxiety interaction reflects the finding that during the third period mean GSR nonspecifics were lower for Lo AACL subjects than for Hi AACL subjects, irrespective of experimental conditions, but that this relationship was not obtained during the first and second periods. This is clearly depicted in Figure 1. Here, during the second experimental period, Hi AACL and Lo AACL subjects in the two experimental conditions showed comparable increases in mean GSR nonspecifics, but during the final rest period, after all subjects had been informed that there would be no more

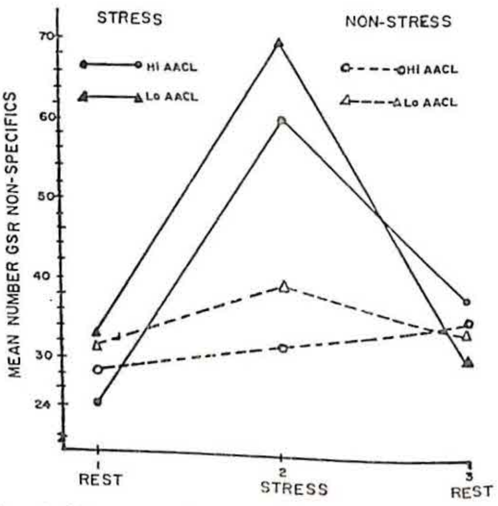


FIG. 1. Mean number of GSR nonspecifics for the four subject groups for the three experimental periods.

threat, their decreases in nonspecifics were not comparable. During the third period GSR nonspecifics decreased less for Hi AACL subjects who had been in the Stress condition than for Lo AACL subjects in the same condition; for subjects who had been in the Nonstress condition, GSR nonspecifics increased for those with Hi AACL while they decreased for those with Lo AACL. Further analyses indicated that for the data of the second and third periods a significant periods by anxiety interaction was obtained for subjects in the Stress condition ($F = 4.48$; $df = 1/24$; $p < .05$), while the obtained periods by anxiety interaction for subjects in the Nonstress condition was not significant ($F = 2.44$; $df = 1/24$). There were no significant periods by anxiety interactions for subjects in the Stress or Nonstress conditions for the data of the first and second periods, or for the first and third periods. Thus it appears that the overall significant periods by anxiety interaction resulted largely from the finding that in the Stress condition Hi AACL subjects showed a smaller mean decrease in GSR nonspecifics after the cessation of threat than did Lo AACL subjects.

DISCUSSION

The present study confirmed the expectation that all subjects exposed to stressful conditions would show greater autonomic arousal than subjects who were not stressed; however, it did not confirm the expectation that stressed subjects with high AACL scores would show greater autonomic arousal than those with low AACL scores. Unexpectedly, it was found that for subjects who were threatened with shock, differential GSR "recovery rates" were observed for Hi AACL and Lo AACL subjects after the cessation of threat.

The present finding that all subjects showed higher mean GSR nonspecifics under stressful conditions is consistent with previous findings (Cohen et al., 1956; Katkin, 1965), and supports the notion that frequency of GSR nonspecific responses is an effective index of emotional response to stress.

The finding that no differences in frequency of GSR nonspecifics were obtained for Hi AACL and Lo AACL subjects during the initial rest period is consistent with the interpretation of the AACL as a measure of transitory or "reactive" anxiety, for on this assumption no differences between high scoring and low scoring groups were expected for the initial rest period. However, the finding that no differences in mean GSR nonspecifics for Hi AACL and Lo AACL groups were obtained during the stress period was not consistent with this interpretation; this finding can be explained only if it is assumed that the threat of shock was sufficiently arousing to evoke greatly increased responses even from relatively nonresponsive subjects. That is, in order to account for the findings of the second experimental period it must be assumed that AACL scores reflect differential tendencies to give autonomic responses to moderately stressful stimuli (see Katkin, 1965). This finding emphasizes the importance of defining the specific stressor situation used to elicit anxiety responses in investigations which evaluate responses to stress (see Endler, Hunt, & Rosenstein, 1962). In order to elicit differential physiological responses to stress from subjects who are presumed to differ in transitory anxiety as measured by the AACL, it would appear that using "personally-involving stress" (Mandler, Mandler, Kremen, & Sholiton, 1961) or "failure stress" (Lazarus, Deese, & Osler, 1952) might be more fruitful than threatening subjects with painful electric shock.

The finding in the present study that after the cessation of stress, Hi AACL subjects in the Stress condition showed a significantly smaller decrease in mean GSR nonspecifics than comparable Lo AACL subjects suggests that the AACL, while unable to differentially predict changes in GSR responses to severe threat, may be an index of a construct which may be termed (for want of a better phrase) "autonomic recovery rate." That is, the AACL appears to reflect the degree to which a subject's autonomic activity returns to prestress levels after the cessation of severe threat. An alternative explanation for this phenomenon might be that after having been fooled by the experimenter's instructions concerning the second period, Hi AACL subjects simply were less willing to believe instructions for the third period than Lo AACL subjects, and consequently showed less recovery. Unfortunately, the present study provides no data on this issue, but it would seem important, in future studies, to interview subjects at the conclusion of the final rest period to ascertain their responses to the instructions given for that period.

In summary, the present study underscores the importance of defining clearly, when investigating the relationship between personality characteristics, stress, and physiological responses, both the nature of specific stressors used to elicit anxiety responses, and the nature of the construct which the personality scale is intended to reflect.

REFERENCES

BENDIG, A. W. Extraversion, neuroticism, and manifest anxiety. *Journal of Consulting Psychology,* 1957, **21**, 398.

CATTELL, R. B., & SCHEIER, I. H. The nature of anxiety: A review of thirteen multi-variate analyses comprising 814 variables. *Psychological Reports,* 1958, **4**, 351–388.

COHEN, S. I., SILVERMAN, A. J., & BURCH, N. R. A technique for the assessment of affect change. *Journal of Nervous and Mental Disease,* 1956, **124**, 4.

ENDLER, N. S., HUNT, J. McV., & ROSENSTEIN, A. J. An S-R inventory of anxiousness. *Psychological Monographs,* 1962, **76**(17, Whole No. 536).

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the Hysteria-Psychasthenia scales from the MMPI. *Journal of Abnormal and Social Psychology,* 1955, **50**, 135–137.

GREINER, T. H., & BURCH, N. R. Response of human GSR to drugs that influence the reticular formation of the brain stem. *Federation Proceedings,* 1955, **14**, 346 (Abstract).

HOYT, D. P., & MAGOON, T. M. A validation study of the Taylor Manifest Anxiety Scale. *Journal of Clinical Psychology,* 1954, **10**, 357–361.

KATKIN, E. S. The relationship between self-report and physiological indices of anxiety during differentially stressful conditions. Unpublished doctoral dissertation, Duke University, 1963.

KATKIN, E. S. The relationship between manifest anxiety and two indices of autonomic response to stress. *Journal of Personality and Social Psychology,* 1965, **2**, 324–333.

LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychological Bulletin,* 1952, **49**, 271–277.

MANDLER, G., MANDLER, JEAN, KREMEN, I., & SHOLITON, R. The response to threat: Relations among verbal and physiological indices. *Psychological Monographs,* 1961, **75**(9, Whole No. 513).

MARTIN, B. The assessment of anxiety by physiological behavioral measures. *Psychological Bulletin,* 1961, **58**, 234–255.

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychological Bulletin,* 1960, **57**, 403–415.

SILVERMAN, R. E. The manifest anxiety scale as a measure of drive. *Journal of Abnormal and Social Psychology,* 1957, **55**, 94–97.

TAYLOR, JANET A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology,* 1953, **48**, 285–290.

WINTER, W. D., FERREIRA, A. J., & RANSOM, R. Two measures of anxiety: A validation. *Journal of Consulting Psychology,* 1963, **27**, 520–524.

ZUCKERMAN, M. The development of an Affect Adjective Check List for the measurement of anxiety. *Journal of Consulting Psychology,* 1960, **24**, 457–462.

ZUCKERMAN, M., & BIASE, D. V. Replication and further data on the validity of the Affect Adjective Check List measure of anxiety. *Journal of Consulting Psychology,* 1962, **26**, 291.

(Received April 5, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 147–150

SOCIAL PARTICIPATION IN HOMOGENEOUS AND HETEROGENEOUS GROUPS OF CHRONIC SCHIZOPHRENICS

HERBERT E. SPOHN[1] AND WILLIAM P. WOLK[2]
Franklin D. Roosevelt Veterans Administration Hospital

In this study the responsiveness of socially withdrawn, regressed chronic schizophrenics to changes in the characteristics of their social environment was examined. It was hypothesized that withdrawn, regressed, hospitalized patients would participate more effectively in problem-solving groups heterogeneous with respect to the activity level and "mental health" of its members than in groups homogeneous in this regard. 16 Ss rated as withdrawn and regressed and 16 active, improved Ss solved group problems in a homogeneous and heterogeneous group context. Ratings by independent judges, employing a scale devised for this purpose, indicated that both withdrawn and active chronic schizophrenics participate significantly more effectively in a heterogeneous group problem-solving situation than in a homogeneous setting.

Observational studies by Schooler and Spohn (1960) and Solomon and Spohn (1958) testify to the marked degree of sustained social withdrawal that characterizes the adjustment of hospitalized chronic schizophrenics. Clinical studies take note of the fact that the range of response to any kind of environmental stimulation in chronic schizophrenics is quite limited, and that such reactions as patients do manifest are simple and stereotyped in nature.

Studies by Schwartz and Stanton (1954) and York, Brown, and Greenblatt (1955) have raised the question of whether the impoverished, stable adjustment of the chronic schizophrenic is autochthonously determined or whether the bland, simplistic homogeneity of the typical hospital-ward environment may not play a role in conditioning the observed adjustment of the chronic schizophrenic. This latter conjecture is strengthened by the observation that large mental hospitals typically assign patients to wards on the basis of the management problem they represent. Similarity of management problems on a ward of patients leads to greater managerial efficiency. Thus, withdrawn, relatively deteriorated chronic schizophrenics are typically grouped together on wards where, in consequence, mutual social stimulation is likely to be minimal. This state of affairs obtained in the hospital where the research to be described was undertaken. The homogeneity of the ward social environment called to mind the literature dealing with problem-solving in small homogeneous and heterogeneous groups. In particular, it had been consistently demonstrated by Hoffman (1959; Hoffman & Maier, 1961) and others that groups heterogeneously composed on various bases are superior in problem-solving tasks to homogeneous groups. It thus suggested itself that socially withdrawn chronic schizophrenics might be stimulated to greater social activity in a group heterogeneous with respect to social withdrawal and mental health. Such a hypothesis if experimentally tested would also throw light on the question of the modifiability of the rigidly stereotyped behavior of the chronic schizophrenic patient and his responsiveness to the social environment, the limits of which the ward environment may not test. Specifically, it was hypothesized in this study that chronic withdrawn schizophrenics participate more effectively in problem-solving groups composed of patients heterogeneous in the extent of withdrawal and apathy than in groups homogeneous in these respects.

METHOD

Subjects

The subjects were 32 hospitalized chronic schizophrenic patients. Sixteen of these subjects had been rated as withdrawn and regressed by clinical psychologists using the Montrose Adjustment Rating Scale (1952) and were under close supervision on closed wards.

---

[1] Now on the staff of the Menninger Foundation in Topeka, Kansas.

[2] Now on the staff of the Brockton VA Hospital in Brockton, Massachusetts.

Sixteen patients had been rated as active and improved and were on privileged status on open wards. The two subject groups were equated for age and education. All subjects had a minimum of 2 years of continuous hospitalization. None of the patients had social contact with each other prior to the experiment.

Procedure

Working in four-man groups, patients were instructed by the experimenter to solve five group problems in succession. The problems were simple, but required participation by all four group members for their solution. For example, subjects were each handed a card with a number on it. They were then instructed to tell the experimenter who among them had the card with the highest number on it. In another task, subjects were instructed to discover, through discussion, two ward activities which all four subjects especially enjoyed. Two sets of five such tasks, equivalent in content and level of difficulty, were developed. Based upon considerable pilot work, these tasks were designed to offer the possibility of participation at some level to the most withdrawn as well as to the most active subjects.[3]

Each subject participated in the problem-solving situation twice, once in a group composed of four subjects of the same level of mental health adjustment (homogeneous group) and once in a group composed of two withdrawn, regressed patients and two active, improved patients (heterogeneous group). The order in which subjects experienced these two types of group composition and performed the two sets of equivalent group tasks was counterbalanced.

The level of social participation of each subject in both types of group composition was rated by four experienced clinical psychologists observing subjects through a one-way vision screen and employing a social participation rating scale especially devised for the purposes of this research. The scale consisted of six major rating categories, namely: organizing (or leadership) behavior; mediating (or independent cooperative) behavior; accommodating behavior (passive cooperation); blocking (passive noncooperative); withdrawing; and interfering behavior. These categories themselves were rated on an 11-point scale by 12 clinical psychologists according to the degree to which each reflected a contribution to the achievement of the group objectives in the type of group tasks employed in the experiment. The median rating for each category was assigned to that category as its numerical weight and in this fashion an "equal-appearing" intervals scale for rating the level of social participation in the group problem-solving situation was obtained.[4]

Independent ratings with this scale by four raters used in the present study on a separate sample of 16 patients were correlated and yielded a composite inter-rater reliability coefficient of .88 (Pearson r).

In each of the two types of group compositions the level of social participation of each subject was rated once for each of the five group tasks. The five ratings thus obtained for each subject in each of the two group compositions were averaged across the four raters, converted to the numerical weights previously derived and summed. This procedure yielded two total scores for each subject representing his level of social participation (as defined by the scaling method employed) in the homogeneous and heterogeneous group situations, respectively.

There is little doubt that raters were aware of differences in the group composition contexts in which they were rating the same subjects on two occasions. The effect of this knowledge upon their ratings was not independently assessed. They were not informed, however, of the hypothesis being tested and there was no evidence that they independently shared the expectations of the experimenters, which might in turn have influenced their ratings. Moreover, as regards this issue of rater bias, it should be pointed out that the specificity and concreteness of behavior descriptions that defined the rating categories (in the use of which raters had been carefully trained) could be expected to vitiate in some measure whatever bias toward hypothesized results might have guided the raters' judgments.

RESULTS

The design of the experiment was factorial, making possible the independent analysis of the effect of four major variables: level of mental health, type of group composition, type of problem-solving tasks, and order of presentation (or practice) upon the level of social participation.

For convenience, type of problem-solving task and order of presentation were analyzed separately. The results of this analysis of variance indicate that significant changes in the level of social participation in this experiment were *not* associated with practice in problem solving, the types of sets of problem-solving tasks administered, or interaction of these two factors. These findings further justified the subsequent independent analysis of the level of mental health and type of group composition variables.[5]

---

[3] Copies of the Social Participation Rating Scale, describing the rating categories, assigned weights and instructions to judges, as well as detailed descriptions of the group problem-solving tasks and instructions have been deposited with the American Documentation Institute. Order Document No. 8714 from ADI Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

[4] It was possible to test the "equivalence" of the two sets of tasks while examining interrater reliability of the social participation of rating scale (to be described below) with a separate sample of 16 subjects. As regards the level of social participation elicited by these two sets of five group problem solving tasks, no differences between them were obtained when the two sets were administered to the same subjects in a counterbalanced order. (Two groups received Set A first and then Set B, to two groups the sets were administered in reverse order.)

It will be noted in Table 1 that the mean social participation scores for the patients in remission are considerably higher than those of the regressed subjects regardless of group composition. On the average, the former function at the level of independent cooperation, whereas the regressed patients on the average function at the level of accommodation, that is, of participating only when patient leadership is available. The level of social participation increases for both groups, however, in the heterogeneous group composition. Analysis of variance, shown in Table 2, indicates that the difference between regressed and improved subjects in level of participation is significant at the .01 level and also that the increase in level of social participation for both groups in the heterogeneous group composition is significant at the .05 level. The difference in the level of social participation between homogeneous and heterogeneous group composition in the regressed group is greater than in the improved group, but in the analysis of variance (see Table 2), the null-hypothesis with respect to the interaction between level of mental health and group composition cannot be rejected.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SOCIAL PARTICIPATION RATINGS OF SUBJECTS IN TWO GROUP COMPOSITIONS

Group composition	Regressed		Improved	
	M	*SD*	*M*	*SD*
Homogeneous	39.5	8.77	57.6	7.14
Heterogeneous	45.8	7.93	60.0	6.24

TABLE 2

ANALYSIS OF VARIANCE: SOCIAL PARTICIPATION RATINGS OF LEVEL OF MENTAL HEALTH AND GROUP COMPOSITION

Source	*df*	*MS*	*F*
Level of mental health[a] (A)	1	4456	66.5**
Group composition (B)	1	248	6.04*
A × B[b]	1	95	2.31
Between subjects in same mental health level	30	67	
B × subjects within health	30	41	

[a] The between subjects in same mental health level source of variance served as the error term to test the two main effects.
[b] The pooled subjects in Health Level × Group Composition Interaction served as the error term to test the Health Level × Group Composition Interaction.
* $p < .05$.
** $p < .01$.

What these results imply is that in the heterogeneous group composition, patients, regardless of level of mental health, tended to participate more effectively than in the homogeneous group composition. This finding confirms the hypothesis and suggests that the postulated phenomenon occurs in active, improved chronic schizophrenics as well. This active group improved its level of participation chiefly by assuming leadership roles vis-a-vis the regressed subjects. (The mean level of participation of the active subjects in the homogeneous situation is within the mediating [independent cooperative] range and shifts into the organizing [or leadership] range in the heterogeneous situation.) The latter improved by accepting such leadership or by responding to the example set by the participating active patients. It is of some interest to note that while all of the regressed patients increased their level of participation, some of the improved patients declined in the heterogeneous situation apparently in response to the apathy of the regressed members of the group.

DISCUSSION

The limited hypothesis this study sought to test—namely, withdrawn, chronic schizophrenics participate more effectively in groups heterogeneous in the social activity level of their members than in groups homogeneous in this regard—is confirmed by the results. A broader issue, however, is raised by this finding. Is heterogeneity of the group composition, per se, effective in raising level of social participation in chronic schizophrenics regardless of characteristic preexperimental activity level, or are the obtained differences principally due to the increased social stimulation that the presence of more active, less withdrawn patients provided for the more withdrawn patients? Since group composition was varied in the design for both active and withdrawn patients it is possible to answer this question. The results indicate that *both* active and withdrawn patients improve in level of social participation in the heterogeneous situation. Thus, while a hypothesis of the effect of heterogeneity per se was not under examination in this study, the results do suggest that level of social stimulation, as such, was not the only variable effective in inducing change in social behavior, but rather that heterogeneity of group composition has generally facilitating effects upon the social behavior of chronic schizophrenics. Had additional groups of normal controls and schizophrenic groups composed of 3 improved and 1 regressed, and 3 regressed and 1 improved patients been included in the design of the study, the results would, however, have been more decisive with respect to the issue of the effects of heterogeneity versus homogeneity of group composition as the critical variable.

---

[5] For both analyses the paradigm for "Analysis of Repeated Measurements of Independent Groups" offered by A. L. Edwards, *Experimental Design in Psychological Research*, New York: Rinehart, 1950, (pp. 288–295) was used.

As regards the broader implications of this study, while only a relatively narrow segment of behavior was examined in this study under artificial conditions, it is felt that the results have a significant bearing upon general questions concerning the nature of the adjustment of chronic schizophrenic individuals. This assumption is consistent with results reported by Wilensky (1959) who found a high positive correlation between the social participation ratings of subjects in the present study and the maturity of their level of perceptual organization as determined by Rorschach performance (DL Score). That is, the behavior segment examined by these findings is shown to be significantly related to a perhaps more central personality variable.

In a sense the present study tested the limits, along a particular dimension, within which variability in the social behavior of chronic schizophrenic patients may be expected. The more general inference that may be drawn from the results of this study is that, within limits, chronic schizophrenics are responsive to changes in the characteristics of the social environment, and that this is the case even for the more regressed and withdrawn patients who, in terms of clinical perspectives, appear to be largely oblivious to their surroundings and who have proven refractory to all concerted treatment efforts.

The results of this study also tend to confirm the impression of many observers that the ward social environment in mental hospitals significantly influences the social behavior of even the most isolated and withdrawn schizophrenic patients. Moreover the present findings are at least consistent with the more general hypothesis that the social environment in which patients function may reinforce or reduce the withdrawal and regression so characteristic of chronic schizophrenic individuals.

REFERENCES

HOFFMAN, L. R. Homogeneity of member personality and its effects on group problem solving. *Journal of Abnormal and Social Psychology*, 1959, 58, 27–32.

HOFFMAN, L. R., & MAIER, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *Journal of Abnormal and Social Psychology*, 1961, 62, 401–407.

RACKOW, L. L., NAPOLI, P., & KLEBANOFF, S. G. A group method for the rapid screening of psychiatric patients. *American Journal of Psychiatry*, 1952–53, 109, 561–566.

SCHWARTZ, M., & STANTON, A. H. *The mental hospital*. New York: Basic Books, 1954.

SCHOOLER, C., & SPOHN, H. E. Social interaction on a ward of chronic schizophrenics. *International Journal of Social Psychiatry*, 1960, 4, 115–119.

SOLOMON, L. F., & SPOHN, H. E. A method for the systematic analysis of social interaction on a ward of chronic schizophrenics. Paper read at American Psychological Association Meeting, Washington, September 1958.

WILENSKY, H. Rorschach developmental level and social participation in chronic schizophrenics. *Journal Projective Techniques*, 1959, 23, 87–92.

YORK, F., BROWN, L., & GREENBLATT, M. *From custodial to therapeutic patient care in mental hospitals*. New York: Russell Sage Foundation, 1955.

(Received April 7, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 2, 151–154

COMMENTS ON THE SO-CALLED ACQUIESCENCE SET IN APHASIA[1]

DAVIS HOWES

Boston University School of Medicine

Some data recently reported as evidence for the hypothesis that aphasic patients exhibit a set to acquiesce are shown not to support that hypothesis when the control data are taken into account. Several factors that need to be considered in research on this type of problem are noted. The broader issues concerning the usage of "yes" and "no" in cases of aphasia, which underlie this research, hold some interesting implications for the development of a theory of language. These phenomena are briefly discussed.

In a recent article, Jenkins and Schuell (1965) offer data from which they conclude that aphasic patients exhibit a set to agree or acquiesce when answering questions put to them. The issue is of theoretical interest beyond its practical implications for the construction of tests for aphasia, because of its indirect bearing on the question whether assertion or negation is better preserved in aphasia and thus more deeply seated in the language mechanism. When their data are carefully examined, however, they are found not to support the notion of an acquiescence set. The reasons are set forth in this paper. Because some of the general issues raised by their research nevertheless remain of interest, certain phenomena of aphasic language that must enter into consideration of those issues are indicated later.

The data supplied by Jenkins and Schuell (1965) consist of the errors made by 30 normal subjects and by more than 50 aphasic patients who answered "yes" or "no" to a series of verbal statements. Sixty statements, divided into four tests of 15 statements each, were used. The differences between these tests need not be considered here; it will suffice to note that half the items in each test called for an affirmative answer ("Is an orange a fruit?") and half called for a negative answer ("Are towns bigger than cities?"). If the subjects commit more errors on the items that call for negative answers, a set for acquiescence is said to exist. This is indeed the case for their aphasic patients, and the difference is shown by statistical test to be reliable. So runs their argument.

The trouble lies in the data from their control group. For convenience let us term an error where a subject says "yes" to a question demanding a negative answer a *false-positive* error, and one where a subject says "no" to a question demanding a positive answer a *false-negative* error. To establish that aphasic patients have an acquiescence set the authors must show, not simply that the patients make more false-positive errors than false-negative errors, but that the excess of errors in this direction is *greater* for patients than for normal subjects. Otherwise there is no basis for asserting that the "acquiescence set" they found in the data from their patients is in any way associated with aphasia. Now, the error rates for their normal subjects, to no one's surprise, are much lower than their rates for their patients: less than a tenth, in fact. Jenkins and Schuell dismissed these data with the remark that their normal subjects showed "virtually no tendency" to make more errors of the false-positive kind.

By dismissing their control data, Jenkins and Schuell have in effect introduced an implicit a priori assumption that on their tests normal subjects give false-positive and false-negative errors in equal proportions. Only on that hypothesis would the significant difference they found for their group of aphasics permit the conclusion they drew. But this assumption can hardly be accepted at face value. It is much easier, generally speaking, to construct difficult items demanding negative answers than equally difficult

---

[1] The preparation of this paper was supported by research grant MH 08472-03 from the National Institutes of Health, United States Public Health Service.

TABLE 1

RATIOS OF FALSE-NEGATIVE TO FALSE-POSITIVE ERRORS IN NORMAL SUBJECTS AND APHASIC PATIENTS

Subject	Test	Error rates		Ratio
		False-Negative	False-Positive	
Normal	I	.000	.024	.000
	II	.013	.033	.394
	III	.010	.004	2.500
	IV	.000	.300	.000
	I–IV	.006	.091	.066
	I–III	.008	.021	.381
Aphasic	I	.069	.303	.228
	II	.111	.199	.558
	III	.074	.140	.544
	IV	.124	.313	.396
	I–IV	.094	.234	.402
	I–III	.085	.209	.406

items demanding positive answers—as the reader may easily satisfy himself by trying his hand at a few. The art of phrasing questions thus contains a built-in bias that tends to produce a spurious appearance of an acquiescence set.

On examination, in fact, their control data do give evidence of this tendency. Pooling the data from all four tests we find that their normal subjects made a total of 87 false-positive errors on the 960 negative items and only 5 false-negative errors on the 840 positive items. The difference between these proportions is significant at the .001 level ($u = 8.47$; using the arc sine transformation to equate variances, $u = 9.65$). In their discussion, Jenkins and Schuell argue that the results for their Test IV, which are the most unfavorable to their hypothesis, are affected by an artifact. If we accept this argument and eliminate the data from that test, we find that the error proportions for the control subjects on Tests I–III are 15/720 for false-positive and only 5/630 for false-negative errors, a difference significant at the .05 level ($u = 2.03$). Thus despite the smaller sample, their control subjects show the same tendency to "acquiesce" that they found with their aphasics.

To justify their conclusion, then, one cannot simply show that a significant difference exists between the number of false-positive and false-negative errors made by the aphasics. Instead, one must test the hypothesis that the excess of false-positive over false-negative errors is greater for the aphasics than for the normals. For this purpose the statistical approach adopted by Jenkins and Schuell is an unfortunate choice. They applied t tests to the differences between the number of false-positive and false-negative errors committed by the patients on each test. To test the appropriate hypothesis in this way it would be necessary to compare these values with the results of tests *of equal power* applied to the control group. Since the power function of the t test depends on the number of observations, it would be necessary to expand their control group until the total number of errors they produced is as great as the number produced by the patients. This would require a control group about 20 times as large as the one they used, to judge from the error rates they report. With a smaller set of observations, a t test is more likely to result in an error of inference of the second kind, in which the value of t falls below the critical level when in fact a true difference between means exists.

These difficulties can be avoided by expressing the data as ratios instead of differences. Let R be the ratio of false-negative to false-positive errors. A tendency to acquiesce is then indicated by $R < 1$. On the hypothesis proposed by Jenkins and Schuell, that an acquiescence set is associated with aphasia, R should be significantly smaller for their aphasic patients than for their normal control subjects. Since the ratios represent unbiased estimates of population values, unlike the differences, they can be compared legitimately despite the greater number of errors made by the patients.

Table 1 presents these ratios calculated from the data supplied by Jenkins and Schuell in their Table 1. For three of their four subtests the ratios are greater for the aphasics than for the normals, contrary to their hypothesis. If the data from all four tests are pooled, the ratios are 0.07 for the control group and 0.40 for the aphasics. If we accept the argument that Test IV is artifactual [2] and pool only the data from Tests I–III, the ratios are 0.38 for the control group and 0.41 for the aphasics, still in the direction contrary to their hypothesis.

Clearly, then, the data reported do not support the notion that aphasics have a greater tendency to acquiesce than normal subjects. Both normal and aphasic subjects, it may be observed, produce more than twice as many false-positive as false-negative errors. This fact bears out the point made earlier, that there is a natural tendency for questions phrased in the positive form to be easier than questions phrased in the negative form.

While their conclusion must be abandoned in the fact of the above analysis, the general topic upon which Jenkins and Schuell focus attention—namely, the use of simple positive and negative utterances in aphasia—is one that contains much that is of interest. These phenomena hold implications that are of some significance for the theory of normal language processes as well as for the evaluation of aphasic disorders themselves.

Very frequently, "yes" and "no" are the first words that the aphasic patient uses appropriately in the early stages of his recovery. Often they remain his only functional words for long periods of time. This in itself is not surprising, since recovery of vocabulary in aphasia generally follows the frequency of occurrence of words (Howes, 1964) and these are among the most frequent words in the language other than connective or interstitial words that have no propositional significance when used in isolation. The concepts that these words denote, moreover, are perhaps the simplest possible from a logical point of view.

But the phenomena associated with the usage of these two little words are far too rich, in normal speech as well as in aphasia, to be encompassed by a purely logical analysis. A delightful article by Critchley (1961) establishes this point with an array of illustrations, as amusing as they are informative, that range widely over the special features presented by "yes" and "no" in various languages and special situations. With regard to the recovery of these words in aphasia, it would be difficult to improve upon the account given by Hughlings Jackson over 80 years ago (Jackson, 1879–1880, reprinted 1958; cf. pages 174–178 of the reprinted edition). Some of these phenomena are worth recapitulating here briefly both for their bearing upon the testing procedures used by Jenkins and Schuell and for the implications they hold for certain current approaches to the theory of language.

It is by no means uncommon for a patient to recover one or the other of these words without the other; or, again, one or the other word may appear as a recurrent or automatic utterance. In such cases the patient is apt to convey his answer to a query calling for a "yes" or "no" answer by extraverbal means. Thus if he has recovered the use of "no" alone, the patient may nod his head emphatically while saying "no" to indicate a positive answer; or he may register his intent by saying "no" with the rising intonation commonly associated with "yes"; or he may repeatedly attempt to say the correct answer, coming out with "no" every time but conveying his meaning unmistakably by an evident display of disgust when the wrong word once more emerges. Other methods, sometimes quite cleverly conceived and often amusing, may be hit upon by a patient to communicate the opposite of what he actually says.

These characteristics of aphasic language would have to be taken into account before one could interpret any positive findings from

---

[2] The argument made by Jenkins and Schuell is that in Test IV the subject faces what is essentially a proofreading task, which is especially conducive to errors of the false-positive kind. This argument, although ad hoc, is plausible as far as it goes; yet it does not fully account for their results on that test. Their aphasics, while making hardly more false-positive errors than the normals on this test, managed to commit more false-negative errors here than on any other test (see Table 1). The ratio of false-negative to false-positive errors for aphasics, in fact, is hardly distinguishable from the ratio based on the other three tests (0.40 compared to 0.41). Since the error proportions are more reliable for the aphasics than for the normals, it may well be that the absence of false-negative errors by the normal subjects on Test IV is only accidental and would disappear with a larger sample. Otherwise we are left with the interesting, but rather startling, conclusion that aphasic patients are much *better* at proofreading, relatively speaking, than are normal subjects. Their argument for ignoring the data from Test IV can therefore be accepted only with reservation.

tests like those reported by Jenkins and Schuell. Suppose, for the moment, that their results had shown a significantly greater proportion of false-positive errors for aphasic subjects compared with normals. A statistically significant difference might result simply if "yes" were a recurrent utterance for one or two of the patients in their series. Yet the full set of facts still would not, in that case, justify a conclusion that there is a set to acquiesce in aphasic patients. The distribution of errors in the different patients is of vital importance in studies of this kind, and in our hypothetical example would suggest the true nature of the result.

An empirical question with interesting theoretical implications, which lies at the heart of the issue posed by Jenkins and Schuell, is whether "yes" or "no" is more often preserved (or first recovered) in aphasia (Critchley, 1961). No data on this point, to my knowledge, have been produced. From a logical standpoint, one might expect the one to occur as often as the other, or perhaps expect "yes," as the simpler concept requiring no change, to predominate. Yet most experienced observers have the general impression that, if there is any difference at all, "no" more often returns first. (It was the discrepancy between this and the Jenkins and Schuell conclusion, in fact, which led to reexamination of their data.) The frequency of "no" in the Lorge Count (Thorndike & Lorge, 1944) is over five times greater than the frequency of "yes"; but these data are for written language, and, as Critchley emphasizes, "yes" and "no" are above all elements of conversational speech rather than written language. Some data on the course of recovery of these primordial words in aphasia might provide a valuable clue to the nature of the processes that govern the embedding of words in the language mechanism.

The manner in which aphasics use and misuse these two words raises a serious question about the relevance of one of the main lines of approach to a theory of language that is followed in psychology. Consider the patient who repeatedly says "no" to a query he wants to answer with "yes," but who communicates his meaning by the disgust he displays when he emits the wrong word again and again. The language mechanism here has been reduced to a rudimentary state without appreciable disturbance of meaning or cognition. Clearly, then, there are at least two physiologically distinct processes at work, whatever names one wishes to assign to them. One therefore can hardly hope to determine the properties of one of those processes from an analysis of the other. Yet this is precisely the course taken by those schools of thought that treat language as a cognitive, symbolic, or semiotic process which can be approached through an analysis of meanings and concepts. Not that those matters are not worth investigation in their own right; but we should not confuse them with the verbal, or language, processes that are damaged in cases of aphasia.

The distinction between these two types of process, although apparent from many of the properties of both normal and aphasic speech, is nowhere more dramatic than in the aphasic patient's use of "yes" and "no." Thus, in its general form, the problem raised by Jenkins and Schuell remains one of considerable interest, despite the negation of their specific conclusion.

REFERENCES

CRITCHLEY, M. The everlasting "no"; the everlasting "yea." *Perspectives in Biology and Medicine*, 1961, **5**, 101–120.

HOWES, DAVIS. Application of the word-frequency concept to aphasia. In A. V. S. de Reuck & Maeve O'Connor (Eds.), *Ciba Foundation Symposium on disorders of language*. London: Churchill, 1964. Pp. 47–75.

JACKSON, J. H. On affections of speech from disease of the brain. In J. Taylor (Ed.), *Selected writings*. Vol. II. New York: Basic Books, 1958.

JENKINS, J. J., & SCHUELL, H. Acquiescence response set in aphasics. *Journal of Abnormal Psychology*, 1965, **70**, 111–113.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

(Received June 15, 1965)

Four New Psychology Texts From Prentice-Hall

THE ABNORMAL PERSONALITY
THROUGH LITERATURE

Edited by Alan A. Stone, McLean Hospital, Massachusetts, and Sue S. Stone. ". . . . (this book) will attempt to present the 'outward forms' of abnormal personality utilizing the portraits drawn in literature. None of these portraits are as stereotyped as the ordinary case histories presented to illustrate abnormal personality types. They are, however, closer to the reality of human nature, which resists oversimplified classification. These examples from fiction hopefully will permit the reader to grasp the traditional classification of psychological disorders as they can be identified within vivid delineation of human lives and human relationship."—*from the Introduction*

January 1966, 423 pp., $4.50

PSYCHOPATHOLOGY OF CHILDHOOD

by Jane W. Kessler, Director, Mental Development Center, Western Reserve University. ". . . obviously written from an attitude of personal experience in clinical practice and teaching, it is richly laced with case study materials. It is not restricted in scope to behavior problems and emotional disturbances, but takes in a somato-psychological, psychosomatic, language, educational and intellectual deficits and difficulties as well."—*from a pre-publication review*

February 1966, 544 pp., $8.25

PERSONALITY RESEARCH:
A BOOK OF READINGS

Edited by Donn Byrne and Marshall L. Hamilton, San Jose State College. Here is a representative picture of current research, theoretical formulations and issues typical of the field of personality—with emphasis on the field as a science. This collection of readings can be successfully used independently or with basic personality texts; it is most ideally suited for use in conjunction with *An Introduction to Personality: A Research Approach,* described below. Includes: articles by Cronbach, Auld, Meehl, Sears, Harlow, MacKinnon, Mednick, Byrne, Taylor, Rogers, and Holt.

March 1966, 432 pp., paperbound, $5.50

AN INTRODUCTION TO PERSONALITY:
A RESEARCH APPROACH

by Donn Byrne, University of Texas. An Introduction to Personality facilitates an understanding of the way in which the scientific method is utilized in studying behavior. Concentrating on current aspects of research which deal with personality as a science rather than a type of speculative philosophy, this new text discusses where ideas originate, how these ideas are translated into research, and how theoretical speculations and research findings are interconnected.

April 1966, 560 pp., $7.75

(PRICES SHOWN ARE FOR STUDENT USE.)

For approval copies, write: Box 903
Prentice-Hall, Englewood Cliffs, N. J. 07632

A new book by the author of
THE PSYCHOLOGY OF AGGRESSION . . .

PSYCHOPATHOLOGY

By ARNOLD H. BUSS, *Professor of Psychology,*
Rutgers, The State University.

PSYCHOPATHOLOGY presents a thorough review of the problems and findings in the field of abnormal psychology. Two principles underlie the organization and content of the book. The first is that exposition should be separated from explanation, data from inferences, and facts from theories. Theories are described with impartiality, and then evaluated and compared. The second principle is that progress may be marked by advances from the clinic to the laboratory, that is, from the clinician's subjective observations to the laboratory researcher's objective, quantitative data. One consequence of both principles is an emphasis on issues, facts, and theories, and the omission of case histories.

The book has four main sections: broad issues, the nature of abnormality, definitions, models, and classifications . . . symptoms and theories of neurosis . . . symptoms and theories of psychosis . . . psychosomatic and conduct disorders.

CONTENTS

Normality and Abnormality
Concepts and Models of Psychopathology
The Domain of Psychopathology
Neurotic Systems
The Psychoanalytic Approaches to Neurosis
Learning Approaches to Neurosis
Comparison of Psychoanalytic and Learning Approaches
Psychotherapy
Symptoms of Psychosis
Dimensions of Psychosis
Regression Theories of Psychosis
Motivational Theories of Schizophrenia
Cognitive Theories of Schizophrenia
Biological Approaches to Psychosis
Interpersonal and Social Approaches to Psychosis
Psychosomatic Disorders

1966. *483 pages.* *Prob. $7.95.*

JOHN WILEY & SONS, Inc., 605 Third Avenue, New York, N. Y. 10016

. . . and other Wiley books of related interest

PSYCHOTHERAPY AND THE PSYCHOLOGY OF BEHAVIOR CHANGE

By ARNOLD P. GOLDSTEIN, *Syracuse University;* KENNETH HELLER, *Indiana University; and* LEE B. SECHREST, *Northwestern University.*

This book develops an approach to psychotherapy which holds that therapeutic efficiency will be most adequately advanced by major attention on the part of both practitioners and researchers to *non-clinical* research findings in contemporary general psychology dealing with behavior change. Drawing upon such materials, the authors present and develop several series of hypotheses suggesting new and major departures in therapeutic techniques and theory. *1966. 472 pages. Prob. $8.95.*

THE INVESTIGATION OF PSYCHOTHERAPY
Commentaries and Readings

By ARNOLD P. GOLDSTEIN *and* SANFORD J. DEAN, *both of Syracuse University.*

A collection of representative, contemporary investigations of psychotherapy, together with critical discussions of the major issues involved in the conduct of psychotherapy research. Throughout the commentaries, an attempt is made to take a fresh look at the issues and problems involved in such research and, in some instances, to recast these into more heuristic terms. All major approaches are represented and recent developments in therapeutic and research methodologies are emphasized. Philosophical approaches and ethical issues are examined; problems of design and measurement are explored in depth. *1966. Approx. 496 pages. Prob. $9.95.*

SYSTEMS OF PSYCHOTHERAPY: A Comparative Study

By DONALD H. FORD *and* HUGH B. URBAN, *both of The Pennsylvania State University.*

"To view a variety of divergent systems in terms of a common framework is a task of no minor proportions. Ford and Urban have made this ambitious attempt and, in this reviewer's opinion, they have done so with considerable skill. This is an important book and constitutes a solid contribution to a field in which controversy abounds and in which there are few firmly established principles. In going about their chosen task the authors do not engage in the axe-grinding that seems so often to occur in this field of concern. They are always critical, but never destructive. To my mind, this is one of the real virtues of this volume."—HANS H. STRUPP, University of North Carolina, in *Contemporary Psychology. 1963. 712 pages. $10.95.*

THE FIELDS OF GROUP PSYCHOTHERAPY

By S. R. SLAVSON. A Science Editions Paperback. *$1.95.*

PSYCHOPATHOLOGY: A Source Book

Edited by C. F. REED, I. E. ALEXANDER, *and* S. S. TOMKINS. A Science Editions Paperback. *$2.95.*

SEX AND BEHAVIOR

Edited by FRANK BEACH, *University of California, Berkeley. 1965. 592 pages. $9.75.*

CLINICAL AND SOCIAL JUDGMENT
The Discrimination of Behavioral Information

By JAMES BIERI, *The University of Texas;* ALVIN L. ATKINS, *Yeshiva University;* ROBIN L. LEAMAN, *New School for Social Research; and* J. SCOTT BRIAR, HENRY MILLER, *and* TONY TRIPODI, *all at the University of California, Berkeley. 1966. 271 pages. $7.95.*

JOHN WILEY & SONS, Inc., 605 Third Avenue, New York, N. Y. 10016

PSYCHOLOGY OF PERSONALITY: Readings in Theory

Edited by WILLIAM S. SAHAKIAN · Suffolk University

Facilitates the study of personality by providing in one source the concise theories of 21 eminent psychologists from Freud through Rogers. The finest works of these major theoreticians have been edited from primary sources exclusively to select the salient portions which present clearly the precise theory of each psychologist. Each chapter is prefaced with a synopsis of the psychologist's predominant attitude in personality theory along with notes on his background, education, and published writings.

1965 · 504 pages · $6.50

PSYCHOTHERAPY RESEARCH:
Selected Readings

Edited by Gary E. Stollak · Indiana University
Bernard G. Guerney, Jr. · Rutgers University
Meyer Rothberg · Eastern Pennsylvania Psychiatric Institute

Covers all areas of psychotherapy research representing the spectrum of therapy variants including individual group, child, adult, neurotic and psychotic. Includes classic studies, recent works, and newest approaches in the research and practice of psychotherapy.

1966

MORALITY AND MENTAL HEALTH

Edited by O. HOBART MOWRER · University of Illinois

Through selected readings, the editor explores "mental illness" in terms of personal morality and individual values. He theorizes that "neuroses" are not only medical problems but also moral ones. Writings of psychologists, clergymen, social scientists, lawyers, and popular authors are included to illustrate that the areas of morality, ethics, values, and religion are undergoing a major transformation. The selections center on the premise that psychopathology must be as concerned with morality per se as with personality disturbance and the connection between the two.

1966 · c. 700 pages

RAND McNALLY & COMPANY
The College Department · Box 7600 · Chicago, Illinois 60680

". . . written with the undergraduate in mind . . . approach is based upon the conviction that many, if not most, of the world's social problems exist because the populace does not possess elementary knowledge of the social sciences."
—from the Preface

An Introduction to Social Psychology

by *Richard Dewey*, University of New Hampshire, and *W. J. Humber*, Consulting Psychologist, Milwaukee

A comprehensive introduction to social psychology, this text is designed to provide the freshman or sophomore student with a clear conception of the nature of human behavior. The authors emphasize a special interaction frame of reference for analyzing human behavior that views personality as the result of the interaction of inherited or physiological traits. After the initial exposure to theory, the student is shown applications of the concepts to personality, personality types, and to various forms of group actions. The third part of the text treats theoretical applications to age groups, from infancy through old age. Actual case studies, drawn from the authors' own professional experiences, illustrate the theories and their applications, as do the numerous photographs and line drawings.

1966, 672 pages, $7.95

Write to the Faculty Service Desk for an Examination Copy

THE MACMILLAN COMPANY, 60 Fifth Avenue, New York 10011

In case you miss the "social" from the former Journal of Abnormal and Social Psychology . . .

JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY

Edited by Daniel Katz

- Basic research and theory in the field of social interaction
- $20.00 a year
- Monthly (2 volumes)

Order from: AMERICAN PSYCHOLOGICAL ASSOCIATION
1200–17th St., N.W., Washington, D.C. 20036

The Journal of SOCIAL ISSUES

spssi

OCTOBER 1964 VOL. XX No. 4

YOUTH AND SOCIAL ACTION

Issue Editors: Jacob R. Fishman and Fredric Solomon

Values and Social Action Analyses: A Preface .... Robert Chin
Youth and Social Action: An Introduction .... Jacob R. Fishman
A Memorandum on Identity and Negro Youth .... Erik H. Erikson
A Matter of Territory .... Robert Coles
Youth and Peace: A Psychosocial Study of Student Peace Demonstrators in Washington, D.C. ....Fredric Solomon
Jacob R. Fishman
The Obedient Rebels: A Study of College Conversions to Conservatism
Lawrence F. Schiff
Individual Patterns in Historical Change: Imagery of Japanese Youth
Robert Jay Lifton
Pro-Social Commitment Among College Students ... Victor A. Gelineau
David Kantor
Biographical Sketches
Abstracts

Annual subscription	$5.50
APA members	4.00
Single members	1.50
Bulk copies (2–9)	1.00
Bulk copies (10–24)	.85

Order from: Journal of Social Issues
P. O. Box 1248
Ann Arbor, Michigan

---

published quarterly by

The Society for the Psychological Study of Social Issues
(a division of the American Psychological Association)

Volume 71 June 1966 Number 3

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Chicago

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute

SEYMOUR SARASON
Yale University

JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR	MARY C. MCGARRAGHY	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Editorial Assistant	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1966 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 155–160

SIZE CONSTANCY IN SCHIZOPHRENIA:

A REANALYSIS[1]

RICHARD H. PRICE AND CHARLES W. ERIKSEN[2]
University of Illinois

Perceptual differences between normals and schizophrenics revealed by traditional size-constancy experiments are questionable because of S variability, procedural difficulties, and problems inherent in traditional psychophysical methods. A review of the literature reveals few consistent results. The present study compared normals, paranoid schizophrenics, and nonparanoid schizophrenics on a size-constancy task which allowed derivation of additional dependent variables from signal-detection analysis and uncertainty analysis. No group differences were revealed using the size-constancy procedure, but signal-detection analysis indicated nonparanoid schizophrenics were significantly lower in perceptual sensitivity. Uncertainty analysis revealed additional response dispositions characteristic of the schizophrenic groups.

Investigators have suggested that the size-constancy experiment reveals basic perceptual differences between normal and schizophrenic populations. However, if such basic perceptual differences exist, they have not yet been adequately demonstrated since all studies are subject to one or more of the following problems.

As Eriksen (1963) has pointed out, individual differences constitute the single largest source of variance in perceptual experiments and it is these differences in less rigorously controlled experiments which tend to correlate with more general personality traits. Differences in the size-constancy behavior of schizophrenics suggested by studies failing to control variables of this kind are equivocal at best. Of particular interest is the problem of reliable psychiatric diagnosis. Silverman (1964) suggests that the paranoid-nonparanoid distinction in particular has utility in reducing individual variability within the schizophrenic population.

Stable estimates of size constancy for a given observer are not likely to be obtained if an insufficient number of experimental trials is provided, and a large error component will be introduced into constancy estimates. Regardless of sample sizes used, group differences are difficult to infer under these circumstances. The psychophysical methods used to obtain the measure of size constancy also present problems. Numerous investigators have pointed out that obtained results in psychophysics are often a function of the procedure used (Goldiamond, 1962). It would appear that difficulties of the type mentioned above produce serious problems in interpreting group differences within a given experiment or in comparing experiments using different psychophysical methods.

Investigators in psychophysics have long been aware of the fact that the "attitude" or "criterion" of the observer will affect his psychophysical judgments (Boring, 1920; George, 1917; Guilford, 1927). As Swets, Tanner, and Birdsall (1961) point out, the traditional methods of psychophysics make provision for only a single free parameter, the sensitivity of the observer. These methods are unable to account adequately for the observer's criterion. If the sensitivity and criterion variables are confounded in traditional psychophysical methods, this presents a fundamental problem in inferring differences in sensitivity between groups through the use of threshold estimates. Swets (1961) argues that the difference between two threshold estimates can be attributed to a change in sensitivity only if it is assumed that the criterion is held constant. Since the threshold provides no independent estimates of sensitivity and criterion, the meaning of a change in the value of an obtained threshold is often ambiguous. Investigations of size constancy in schizophrenia are especially vulnerable to these criticisms since one would expect criterial or attitudinal components of the perceptual response to be especially strong sources of variance.

---

[1] This investigation was supported by a Public Health Service Research Career Program Award K6-MH-22,014.

[2] The authors wish to thank Roy M. Hamlin and the staff of the Veterans Administration Hospital, Danville, Illinois, for their cooperation in providing research facilities.

Studies of size constancy in schizophrenia fall into two general categories with respect to the sample studied: (*a*) experiments which compare a schizophrenic sample of unspecified type with a normal control group; (*b*) experiments which compare normals with various subclassifications of the schizophrenic population.

Investigations of the first type show little in the way of consistent findings. Results are approximately equally divided between those which indicate that schizophrenics are overconstant (Maes, 1957; Perez, 1961; Sanders & Pacht, 1952), underconstant (Crookes, 1957; Hamilton, 1963; Weckowicz, 1957), or not different when compared with normals (Leibowitz & Pishkin, 1961; Pishkin, Smith, & Leibowitz, 1962). Investigations of the second type, which examine various subclassifications of schizophrenic patients are somewhat better with respect to consistent findings. Weckowicz and Blewett (1959) found that hebephrenic schizophrenics show underconstancy while paranoids show overconstancy. Rausch (1952) reported that normals and nonparanoid schizophrenics do not differ in constancy but that paranoid schizophrenics show overconstancy. Hartman (1962) confirmed Rausch's findings using afterimage observations. As Harway and Salzman (1964) point out, findings of whatever type are rationalized with post hoc explanations. Such dubious after-the-fact explanations do little to advance knowledge of differences which may exist in the perceptual behavior of schizophrenic subjects.

It is suggested that alternative methods of analysis can be employed which have utility for revealing perceptual differences which may exist between schizophrenics and normals. A method of analysis which allows estimates of sensory capabilities independent of subjective criteria used by the subject in making perceptual judgments is derived from the theory of signal detection (TSD) discussed by Swets et al. (1961). A second method, uncertainty analysis, which describes the subject's response variability and accuracy of stimulus identification, has been described by Garner and Hake (1951), McGill (1954), and Attneave (1959), and is derived from the general context of information theory. These two methods provide a number of dependent variables which, if used in addition to the traditional point of subjective equality (PSE) measures in the size-constancy experiment, provide a more comprehensive description of the perceptual behavior of the subject. Thus an experiment designed to understand perceptual behavior of schizophrenic groups within the size-constancy paradigm would allow for data analysis utilizing several different approaches including (*a*) the traditional size-constancy analysis, (*b*) an analysis providing an estimate of perceptual sensitivity independent of criterial variation, and finally, (*c*) an analysis providing a description of response variability and accuracy of stimulus identification.

METHOD

Subjects. The subjects in the paranoid and nonparanoid schizophrenic groups were selected from the Veterans Administration Hospital, Danville, Illinois, and met the following selection criteria: (*a*) male, (*b*) age range of 25–45 years, (*c*) visual acuity 20/25 corrected or uncorrected, (*d*) no neurological involvement, (*e*) no history of insulin or electroshock, (*f*) open ward, (*g*) consistent diagnosis, (*h*) 1 year or less present hospitalization, (*i*) brief screening interview to assess cooperation, (*j*) comparable educational level. The subjects in the control group were hospital employees who had no history of psychiatric disturbance and met all the above relevant criteria. There were 15 subjects in each of the three groups.

Apparatus. The viewing alley was a stimulus surface painted deck grey, 48 × 96 inches, mounted 45 inches from the floor. The surface was enclosed by two ends 40 × 48 inches. A reduction tunnel was mounted at the subject's end of the viewing alley with a light-tight binocular eyepiece which extended 8 inches beyond the end of the viewing surface on the outside and 12.5 inches inside, with an aperture 5 inches wide and 3 inches high. The viewing surface was illuminated by two 40-watt Amplex Cool White fluorescent bulbs mounted 38 inches above the surface. Twelve white cardboard reflectors 8 × 12 inches were placed 8 inches apart along the light source providing uniform illumination over the entire viewing surface of 17.0 apparent footcandles.

The stimuli consisted of cardboard squares painted flat black. The standard stimulus (*St*) was 30/32 inches on each side. The 10 comparison stimuli (*Co*) ranged from 25/32 inches to 29/32 inches, and from 31/32 inches to 35/32 inches in $\frac{1}{32}$-inch steps. The distance of the *St* from the eye of the subject was 58 inches and always appeared on the right of the subject's visual field. The distance of the *Co* from the eye of the subject was 50 inches and always appeared on the left of the subject's visual field. The visual angle subsumed by the *St* was .9256 degrees. The visual angles subsumed by the 10 *Co* were, from smallest to largest: .8948, .9298, .9648, 1.000, 1.0350, 1.1050, 1.1400, 1.1750, 1.2100, 1.2450 degrees.

Presentation procedure. The following instructions were read to the subject while he simultaneously read a copy handed to him:

> When you look into the viewer you will see two black squares. The idea of the task is to tell me whether the square on the left is really larger or smaller than the square on the right. The square on the right will always be the same size. One half of the squares that you see on the left will be larger and one half will be smaller than the square on the right. Each time you are shown two squares, you are to tell me whether the square on the left is really larger or smaller than the square on the right. Some of the squares will be easy to judge and some will be more difficult. No one is able to judge them all correctly so just do your best. After you have said whether you think the square on the left is larger or smaller than the one on the right, I want you to tell me how confident you are of your judgment. Tell me whether you were positive, fairly sure, or whether it was a guess. For example if you are fairly sure that the square on the left was larger, you would say, larger, fairly sure. Try to use all three of the ratings, positive, fairly sure, and guess.

A series of 80 practice trials were presented to the subject with the *Co* presented in a random order. Following this, 200 experimental trials were run with each *Co* presented 20 times in a random order. Between trials the viewing surface was hidden from the subject's view by a shutter built into the reduction tunnel. The subjects were allowed as much time as they required to respond on each trial but response latency was typically less than 5 seconds.

The instructions required a judgment of the *Co* as larger or smaller, plus a confidence rating of positive, fairly sure, or guess, and thus yielded a rating scale taking the following form: (*a*) larger, positive, (*b*) larger, fairly sure, (*c*) larger, guess, (*d*) smaller, guess, (*e*) smaller, fairly sure, and (*f*) smaller, positive.

Perceptual measures. A PSE was obtained for each subject using the method of constant stimuli (Woodworth & Schlosberg, 1954). Operating characteristics were obtained for each subject using the rating method (Egan, Schulman, & Greenberg, 1959). The hit rate, $P(J_L|S_L)$, is the probability that the subject will judge the *Co* as larger when it is, in fact, larger. The false alarm rate, $P(J_L|S_S)$, is the probability that the subject will judge the *Co* as larger when it is smaller. These values were computed for each of the criteria held by the subject described by each of the categories of scale. Since the a priori probabilities of *Co* being larger or smaller were equal, it was possible to obtain the perceptual sensitivity value of d_s, a measure of perceptual sensitivity or confidence, independent of the subject's criterion or confidence, as described by Egan, Greenberg, and Schulman (1961). Measures obtained in the uncertainty analysis were derived by the method described by Attneave (1959), Garner and Hake (1951), and McGill (1954).

TABLE 1

MEANS AND STANDARD DEVIATIONS OF AGE AND EDUCATION FOR NORMALS, PARANOID SCHIZOPHRENICS, AND NONPARANOID SCHIZOPHRENICS

	Normal	Paranoid schizophrenic	Nonparanoid schizophrenic
Age			
M	35.07	36.93	34.93
SD	6.54	5.08	5.12
Education			
M	11.27	10.28	10.53
SD	2.06	2.29	1.48

RESULTS

Equivalence of groups. The subjects were matched on all selection criteria mentioned above. Table 1 shows the mean and standard deviation for the three groups' age and education. To insure equivalence of groups with respect to these two variables, one-way analyses of variance were performed and no significant differences were found for age or education ($F < 1$).

Size-constancy analysis. A PSE was obtained for each subject. The mean PSE for the control group was 28.87, for the paranoid schizophrenics, 28.74, and for the nonparanoid schizophrenics, 29.47. A one-way analysis of variance indicated no significant differences between groups ($F < 1$). Since a PSE of 30.0 would indicate perfect size constancy, it is evident that all three groups were slightly under-constant in performance. This is explainable in terms of the reduction of distance cues provided by the use of a reduction tunnel and by the homogeneous illumination of the viewing surface. Thus, the traditional size-constancy experiment using PSE

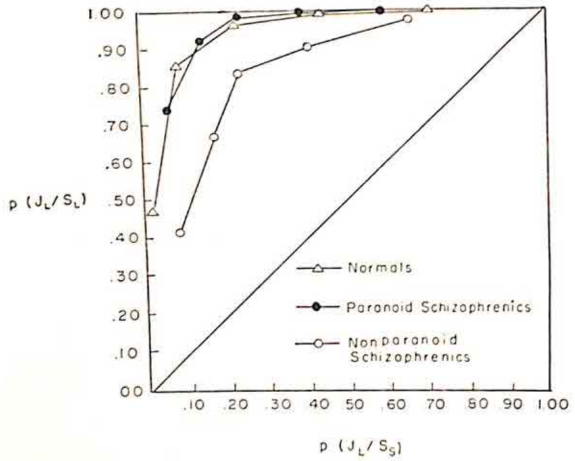


FIG. 1. Averaged operating characteristic curves for schizophrenic and normal groups.

as a dependent variable did not reveal group differences.

Signal-detection analysis. A measure of perceptual sensitivity, d_s, was computed for each subject. The larger the value of d_s the greater the sensitivity. A one-way analysis of variance indicated significant differences between groups—F (2, 42) = 23.73, $p < .001$. The mean values of d_s were 3.184 for the normals, 2.936 for the paranoid schizophrenics, and 1.724 for the nonparanoid schizophrenics. Multiple comparisons of group means were performed which indicated no difference between normals and paranoid schizophrenics, but that the nonparanoid schizophrenics were significantly lower with respect to d_s ($p < .01$).

Average receiver operating characteristic curves (ROC) were derived for each group by averaging hit rate and false-alarm rate over groups for each of the six response categories or confidence ratings. Figure 1 shows the group average ROC curve plotted on cumulative probability coordinates.

It is evident that the ROC curves for normal and paranoid schizophrenic groups are very similar but that the curve for nonparanoid schizophrenics shows markedly poorer performance. Thus the d_s independent of the subject's criterion provides information concerning the performance of the three groups not obtained using the PSE.

Uncertainty analysis. The performance of each subject was analyzed using uncertainty analysis (Attneave, 1959) since this method provides dependent variables which further describe the perceptual behavior of the subjects. For each subject a matrix was constructed along one axis of which were the 10 different sizes of the *Co* stimuli and along the other axis the six response categories from "larger positive" to "smaller positive." In the cells of the matrix were the frequencies with which the subject had responded to each of the 10 stimuli with the specific response categories. From these matrices two statistics were computed. The first of these was $H_X(Y)$ which is a measure in bits of the uncertainty of the response given a particular stimulus. It reflects the predictability of the subject's response if one knows which stimulus he is judging. A large value for $H_X(Y)$ would indicate that the subject was inconsistent or quite variable in which of the six rating categories he applied to the given stimulus on different trials. The second statistic, $H_Y(X)$, is a measure in bits of the uncertainty of the stimulus given the response and thus reflects the predictability of which stimulus had occurred if we know what response category the subject has used.

The mean values of response uncertainty were 1.15, 1.08, and 1.52 bits for the normals, paranoid schizophrenics, and the nonparanoid schizophrenics, respectively. A one-way analysis of variance indicated significant differences between groups—F (2, 42) = 28.78, $p < .001$. Multiple comparisons of means showed the nonparanoid group to be significantly higher than the other two groups with respect to $H_X(Y)$, response uncertainty ($p < .05$).

This analysis revealing a difference between the nonparanoid schizophrenics and the paranoid schizophrenics and normals is consistent with the group differences observed on the d_s statistic. This latter statistic had indicated that the nonparanoid schizophrenics were less sensitive to size-distance discriminations than were the other two groups and the present analysis reveals that their poorer sensitivity was due, at least in part, to their inability to apply the same response category to a given *Co* when it reoccurred.

While the nonparanoid schizophrenics have as a group greater response uncertainty, an analysis of stimulus uncertainty, $H_Y(X)$, provides further information about differences in judgmental behavior among the three groups. If we know which of the six response categories the subject has used on a given trial and we are attempting to predict which stimulus he judged, our uncertainty in bits is 1.96 if the subject was a member of the normal group, 2.26 if he was a member of the paranoid group, and 2.43 if a member of the nonparanoid group—$F\ (2, 42) = 9.86$, $p < .01$. Multiple comparisons of these group means indicated that the normal group was significantly lower in $H_Y(X)$ than the two schizophrenic groups ($p < .05$).

The greater stimulus uncertainty for the nonparanoid schizophrenics is not surprising since it would appear again to reflect their lack of sensitivity and general confusion in size-distance discriminations. However, the difference between the paranoid schizophrenics and normals on stimulus uncertainty requires further examination since these two groups did not differ significantly nor appreciably in sensitivity in detecting size-distance cues. With comparable sensitivity the differences between these two groups on stimulus uncertainty suggests that the paranoids are different from the normals in their use of the confidence-rating categories. In fact, it was quite apparent while running the subjects that the paranoids tended to be predominately "positive" in their judgments. Once they had decided that a *Co* was larger or smaller, they tended to be positive about the judgment irrespective of the amount of physical size difference upon which the judgment was based. Thus, while the paranoids were about equally accurate with the normals in judging whether a *Co* was larger or smaller than the standard, they differed from the normals in having much greater subjective confidence in their judgments.

This interpretation of the difference between the paranoid schizophrenics and normal groups on stimulus uncertainty is supported if the three groups of subjects are compared on the number of times the "positive" confidence rating was used for a judgment. The mean number of positive ratings out of a total of 200 experimental trials was 76.20 for the normals, 120.93 for the paranoid schizophrenics, and 97.13 for the nonparanoid schizophrenics. A one-way analysis of variance indicated significant differences between groups—$F\ (2, 42) = 9.87$, $p < .01$. A multiple comparison of group means revealed that the paranoid group used the positive ratings significantly more often ($p < .05$) than did either the nonparanoid schizophrenics or the normals.

DISCUSSION

The results reported above may be summarized as showing first, that the traditional psychophysical method of PSE fails to differentiate the paranoid schizophrenics, the nonparanoid schizophrenics, and the normals on size-distance discrimination. Second, modern, more analytical methods of studying judgmental processes are capable of detecting differences between these three groups. In particular, the methodology derived from TSD and information analysis reveals that normals and paranoid schizophrenics are comparable in size-distance sensitivity whereas nonparanoid schizophrenics are much less sensitive. Further, the information analysis reveals that the nonparanoid schizophrenics are less sensitive to size-distance cues because of confusion or lack of reliability in their responses from trial to trial. Also the paranoid schizophrenics are found to differ from the normal subjects in the assignment of subjective confidence to their judgmental responses.

These differences in the judgmental process between these three groups can be readily masked if insensitive nonanalytic psychophysical procedures are employed. The failure to find differences between the three groups when compared on the PSE is attributable to the fact that this psychophysical measure is a global statistic, insensitive to variance or unreliability of response, and confounded with the subject's subjective criterion. Three important characteristics of the judgmental process can be easily identified. First of all, there is a sensitivity to changes in the physical stimulus dimension. Second, is the subject's criterion of how confident he has to be in detecting a difference before he will report the difference exists. And third, there is the question of reliability or consistency with which subjects can use responses over trials. The PSE is simply a statistical average describing the stimulus value at which one half of the comparison stimuli are judged as larger and one half are judged as smaller. It is not sensitive to subject confusion of comparison stimuli and thus obscures any existing group differences in stimulus confusion.

The difference between the paranoid schizophrenics and normals in the use of confidence ratings is of particular interest. These two groups were comparable on the d_s which suggests that the paranoids are no more sensitive or accurate in detecting size-distance cues than are the normals. But once the paranoid schizophrenic subject has made the discrimination or judgment that a given comparison stimulus is larger or smaller, his subjective confidence in his judgment is not as highly correlated with the strength of the cue as is that of the normal subject. Even when the cues are too small for reliable discrimination the paranoid schizophrenic tends to ignore the unreliability of judgments in this range and reflects an unrealistic subjective confidence in his judgment. This characteristic of paranoia would seem to fit well with many of the clinical descriptions of this syndrome.

REFERENCES

ATTNEAVE, F. *Applications of information theory to psychology*. New York: Henry Holt, 1959.

BORING, E. G. The control of attitude in psychophysical experiments. *Psychological Review,* 1920, **27**, 440.

CROOKES, T. G. Size constancy and literalness in the Rorschach Test. *British Journal of Medical Psychology,* 1957, **30**, 99–106.

EGAN, J. P., GREENBERG, G. Z., & SCHULMAN, A. I. Interval of time uncertainty in auditory detection. *Journal of the Acoustical Society of America,* 1961, **33**, 771–778.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America,* 1959, **31**, 768–773.

ERIKSEN, C. W. Perception and personality. In J. M. Heine & R. W. Wepman (Eds.), *Concepts of personality*. Chicago: Aldine, 1963. Pp. 31–62.

GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychological Review,* 1951, **58**, 446–459.

GEORGE, S. S. Attitude in relation to the psychophysical judgment. *American Journal of Psychology,* 1917, **28**, 1–37.

GOLDIAMOND, I. Perception. In A. Bachrach (Ed.), *Experimental foundations of clinical psychology*. New York: Basic Books, 1962. Pp. 280–340.

GUILFORD, J. P. Fluctuations of attention with weak visual stimuli. *American Journal of Psychology,* 1927, **38**, 534–583.

HAMILTON, V. Size constancy and cue responsiveness in psychosis. *British Journal of Psychology,* 1963, **54**, 25–29.

HARTMAN, A. M. The apparent size of after-images in delusional and nondelusional schizophrenics. *American Journal of Psychology,* 1962, **75**, 587–595.

HARWAY, N. I., & SALZMAN, L. F. Size constancy in psychopathology. *Journal of Abnormal and Social Psychology,* 1964, **69**, 606–612.

LEIBOWITZ, H. W., & PISHKIN, V. Perceptual size constancy in chronic schizophrenia. *Journal of Consulting Psychology,* 1961, **25**, 196–199.

MAES, J. T. Size constancy in schizophrenia. Unpublished manuscript, Michigan State University, 1957.

McGILL, W. J. Multivariate information transmission. *Psychometrika,* 1954, **19**, 97–116.

PEREZ, P. Size constancy in normals and schizophrenics. In W. H. Ittelson & S. B. Kutash (Eds.), *Perceptual changes in psychopathology.* New Brunswick: Rutgers University Press, 1961. Pp. 39–55.

PISHKIN, V., SMITH, T. E., & LEIBOWITZ, H. W. The influence of symbolic stimulus value on perceived size in chronic schizophrenia. *Journal of Consulting Psychology,* 1962, **26**, 323–330.

RAUSCH, H. L. Perceptual constancy in schizophrenia. *Journal of Personality,* 1952, **21**, 176–187.

SANDERS, R. S., & PACHT, A. R. Perceptual size constancy of known clinical groups. *Journal of Consulting Psychology,* 1952, **16**, 440–444.

SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review,* 1964, **71**, 352–379.

SWETS, J. A. Is there a sensory threshold? *Science,* 1961, **134**, 168–177.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review,* 1961, **68**, 301–340.

WECKOWICZ, T. E. Size constancy in schizophrenic patients. *Journal of Mental Science,* 1957, **103**, 475–486.

WECKOWICZ, T. E., & BLEWETT, D. B. Size constancy and abstract thinking in schizophrenic patients. *Journal of Mental Science,* 1959, **105**, 909–934.

WOODWORTH, R., & SCHLOSBERG, H. *Experimental psychology*. New York: Henry Holt, 1954.

(Received November 1, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 161–164

TREATMENT OF INSOMNIA USING A VARIANT OF SYSTEMATIC DESENSITIZATION:

A CASE REPORT

JAMES H. GEER AND EDWARD S. KATKIN
University of Pennsylvania *State University of New York at Buffalo*

This report describes the treatment of a case of severe insomnia. The treatment, a variant of systematic desensitization, was short and apparently quite successful. The primary difference between usual desensitization procedures and those employed in this case was the absence of a fear hierarchy. A trainee-observer was present during several training and early therapy sessions. He observed 2 therapy sessions and then became the therapist. He was directly supervised for 2 more therapy sessions and then continued treatment singly. The case illustrates the flexibility of behavior therapy in both procedural variations and in training of therapists.

The treatment of behavior disorders by techniques derived from learning theory has been receiving increased attention in current psychological literature (e.g., Eysenck, 1960; Grossberg, 1964; Wolpe, Salter, & Reyna, 1964). Among these techniques, Wolpe's (1958) method of systematic desensitization has been among the most widely reviewed. Critics of systematic desensitization, while often acceding to the efficiency of this technique for the treatment of phobic disorders, claim that it suffers from being limited *solely* to the treatment of phobias. The major purpose of this paper is to describe a case history in which a modification of systematic desensitization was employed successfully in the treatment of a nonphobic disorder—insomnia.

Although insomnia has long been known to clinicians and their patients as a common symptom of behavior disorders, the literature concerning this problem is surprisingly sparse, and consists primarily of theoretical formulations, frequently from the psychoanalytic viewpoint. For example, Fenichel (1945) says of insomnia that

> the cathexes of repressed wishes . . . make sleep impossible. Also acute worries or affect-laden expectations, whether agreeable or disagreeable, particularly sexual excitement without gratification, make for sleeplessness. In the case of neurotic disturbances of sleep, the unconscious factors of course outweigh the others [p. 189].

Discussions of the treatment of insomnia have usually been limited to descriptions of the relative merits of drugs (McGraw & Oliven, 1959) or hypnosis (Wolberg, 1954). Psychotherapeutic intervention is usually directed to the psychodynamic basis presumed to underlie the symptom, and rarely to the symptom per se.

A second purpose of this paper is to demonstrate the unique possibilities that systematic desensitization offers for the training of psychotherapists. Geer (1964) has noted that trainees may observe therapy sessions without appreciably disturbing the course of treatment. In the present case, an observer in the early stages of therapy actually became the therapist in later stages. Thus, not only was a trainee able to observe treatment at first hand, but in addition he was able to become the therapist and subsequently be observed directly by a supervisor.

CASE HISTORY

The client (Miss H.), a 29-year-old, female, white, single, Roman Catholic, was self-referred to the Psychological Clinic of the State University of New York at Buffalo. Her complaint was severe insomnia. She reported that since she terminated her engagement some 12 months before coming to the clinic she had been suffering from sleeplessness. The client reported that she would typically have great difficulty in sleeping for 5 or 6 nights consecutively; then there would usually be 1 night of 6–7 hours of sleep, presumably in response to physical exhaustion, followed by another week of sleeplessness. Miss H. stated that this general sleep pattern had been persistent for the past 12 months. The client was a registered nurse, who had recently returned to college for a BS degree, and she felt that her insomnia was interfering markedly with school performance as well as with most other aspects of her daily life. She felt that she was generally quite irritable as a result of her extensive sleep loss.

Clinical interviewing revealed no evidence of severe psychopathology, nor the presence of any symptoms other than insomnia. An MMPI was administered, and it too was interpreted as revealing no evidence for severe psychopathology; all of the clinical scales were well within the normal range, and the validity scales showed no evidence of dissimulation.[1] Attempts to fit the client into any of the traditional nosological categories were difficult; thus no diagnostic label was attached to her.

Two factors suggested that this case be accepted for behavior therapy. First, the client reported that in 6-weeks time she would be leaving on an extended vacation; thus it appeared that traditional psychotherapeutic procedures would not be feasible because of time limitations. Second, it was felt that a specific symptom reaction such as hers, with its relationship to excessive tension, would be amenable to systematic desensitization. The client was fully informed about the exploratory nature of the procedures to be used, and about the use of the case as a training situation. She consented to enter treatment under these conditions since her insomnia posed an immediate and severe problem from which she desired relief.

COURSE OF TREATMENT

A total of 14 sessions were conducted with Miss H. The first session—the initial interview—was conducted by both authors. At this time information concerning the details and background of the problem was gathered. Also, an attempt was made to assess the suitability of Miss H. for the type of treatment considered. Miss H. reported that three specific themes recurred during her sleepless nights. One concerned her former fiancé; she reported that she still felt an attachment to him and she ruminated over the implications and meaning of her decision to terminate their engagement. The second disturbing theme was concern for the future; for example, Miss H. would wonder if she should continue in school or return to full-time nursing. The third disturbing theme which filled her sleepless hours was concern over her inability to fall asleep. Miss H. would find that not falling asleep would start a vicious cycle in which concerns over falling asleep led to increased sleeplessness and so on. Apparently these three disturbing themes occurred with equal frequency, and appeared to be equally effective in eliciting negative affect.

Four training sessions followed the intake session. These were similar to those described by Geer (1964) in the treatment of a phobia. During the training sessions the client was instructed in relaxation, rehearsed in visual imagery, and asked to construct a hierarchy relating to insomnia (see Geer, 1964; Wolpe, 1958). With respect to this last point, one major problem became obvious; the client could not construct a fear hierarchy related to her insomnia. She reported that she did not feel anxious when anticipating going to sleep, nor did she report anxiety when preparing to retire. Her description of her feelings was that only after retiring would her mind begin to race with upsetting thoughts. It was decided to proceed with treatment, using the presentation of a single item for visualization since construction of a fear hierarchy, traditionally used in systematic desensitization, was impossible.

During the first training session the client became markedly upset. While attempting to relax in a reclining chair Miss H. became agitated, fearful, and tense. She sat up and said that she could not relax with both the therapist and the observer in the room. After some discussion, it was decided that the trainee should leave, and he did. Even then, however, the client remained unable to relax, so the session was terminated. At the second training session the client asked that the trainee not be present; thus, this session was conducted by one therapist only. Miss H. complained that ever since the breakup with her fiancé she had been unable to relax around men. After about 10 minutes of discussion of her concerns surrounding the therapy situation she decided to try to relax once again. This attempt was only moderately successful; however, the client did not become as upset on this occasion. The third training session did not evoke any emotional upset. On the fourth and final training session the trainee returned; Miss H. did not become upset during this session.

---

[1] The client's MMPI was coded by the Hathaway (1956) system as follows: '49$\underline{87}$—$\underline{136}$ (51) 2:6:13. This configuration has been associated with insensitive, emotionally immature patients, who encounter great difficulty in establishing enduring interpersonal relations (see Dahlstrom & Welsh, 1960). Such patients are frequently seen as poor risks for psychotherapy.

Nine therapy sessions directly followed the training sessions and were conducted twice a week. These sessions generally consisted of a brief interview concerning the sleeping pattern of the previous few days, followed by relaxation and the presentation of a single item to visualize—that of the client at home in bed, trying to fall asleep. This single item was presented throughout the therapy sessions with the following modifications: first, Miss H. was instructed to visualize herself lying in bed; then she was instructed to imagine that her mind was racing and that she was unable to fall asleep; finally, she was told to turn her attention away from her disturbing thoughts and to think only of relaxation. At the end of this sequence the client responded, with a finger movement, to questions asking if the scene was clear, if she was able to experience the discomfort, and if she was successful in returning her attention to relaxation. As the therapy sessions progressed, Miss H. reported becoming increasingly able to perform these tasks.

After the trainee had observed two therapy sessions it was felt that he could act as therapist. Therefore, in the third therapy session the trainee became the therapist, and the original therapist became the observer-supervisor. If any supervisory comments became necessary, hand-written notes were exchanged. The client, whose eyes were closed, was unaware of these exchanges. After two supervised sessions both therapists agreed that the trainee was able to continue the procedure alone. Thus, after the fourth therapy session the trainee continued the procedures without an observer.

For the first six treatment sessions the course of therapy was uneventful. Miss H. reported no change in her sleeping pattern and was still quite disturbed by insomnia. However, at the beginning of the seventh therapy session the client reported some change in her sleeping pattern. She reported sleeping more during the period between the sixth and seventh sessions. At the beginning of the eighth therapy session Miss H. gave a similar report, and at the beginning of the ninth therapy session she reported that she had slept well each of the previous nights. Because of final exam pressures and preparations required for her vacation trip, the client decided to terminate treatment at the end of the ninth therapy session. She was asked to return just prior to leaving on her vacation. At that time, 1½ weeks after the final therapy session, she reported a continuation of good sleep. An interview by both therapists revealed no evidence of symptom substitution.

The next contact with the client was approximately 1½ months following termination. At that time, she noted on a postcard from her trip that her sleeping problem had not returned, and that she was "having a wonderful time." A final follow-up interview was conducted 8 months after the final therapy session. At that time the client reported being quite satisfied with her sleeping pattern. She said that she had occasional nights when sleeping was delayed 2–3 hours; however, these nights never occurred more than once a week and often occurred only once in 2 weeks. She also noted that the sleepless nights did not fall consecutively and that they were not anxiety provoking—that is, she did not worry or experience anxiety even though she might not fall asleep for several hours. Miss H. also reported that she felt her general disposition had improved as a result of reduced fatigue. Her response to the question, "How would you characterize the way it used to be with the way you are now?" was, "I had a constant feeling of anxiety and being upset and not being able to think clearly. Now . . . it is entirely different now. I feel relaxed and I'm in control of the situation. I'm more cheerful now." A detailed and careful inquiry was unable to uncover any evidence for symptom substitution. Indeed, there has been a positive spread of therapeutic effect in that her general feeling of well-being has improved as a function of regular sleeping.

DISCUSSION

This case demonstrates several interesting points relevant to behavior therapy. First is the extension of a behavioral technique to a class of disorders usually considered in the province of more traditional techniques. With respect to this point, it is important to note that Jacobson's (1938) classic work with the technique of "progressive relaxation" bears some resemblance to the techniques employed in the present case. With specific reference to the treatment of insomnia, Jacobson (1938) noted: "Years of observation on myself suggested in 1910 that insomnia is always accompanied by a sense of residual tension [2] and can always be overcome when one successfully ceases to contract the parts in this slight measure [p. 29]." In discussing the general psychotherapeutic advantages of progressive relaxation Jacobson noted: "Accordingly, present results indicate that an emotional state fails to exist in the presence of complete relaxation of the peripheral parts involved [p. 218]." It should be noted also that the techniques employed in the present case are different from Jacobson's in some respects. Jacobson did not conceptualize the insomniac's difficulties in terms of tension associated with specific conditioned fear responses. Furthermore, Jacobson's technique required that the patient practice relaxation over a much wider range of muscle groups, and for a considerably longer period of time; some of his patients practiced relaxation at home for 1–2 hours per day. Finally Jacobson did not employ the technique of visualization. It is possible, of course, that visualization of the scenes was irrelevant, and that the client simply learned to relax; however, it appears that this alternative hypothesis is testable, and should be subjected to empirical investigation.

A second point of interest is the unique training opportunity offered by behavior therapy. In this case the trainee was able to observe at first hand the techniques employed, and then take over the procedures without a serious setback in progress. In fact, the finding that only nine therapy sessions were required suggests that if the transfer was detrimental, the effects were transient. Obviously, only controlled studies can adequately answer this question. The successful transfer of the client from one therapist to another supports the suggestion that because of its reduced emphasis upon the patient-therapist relationship behavior therapy is relatively "therapist free" as compared with usual approaches to psychotherapy.

This paper also demonstrates the flexibility of behavior therapy. In the present case no fear hierarchy was utilized; thus it appears that in appropriate circumstances a hierarchy is not necessary. In summary, this case illustrates the successful use of a variant of systematic desensitization both for the treatment of a common symptom of behavior disorders and for the training of a psychotherapist.

REFERENCES

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook.* Minneapolis: Univer. Minnesota Press, 1960.

EYSENCK, H. J. (Ed.) *Behavior therapy and the neuroses.* New York: Pergamon, 1960.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: W. W. Norton, 1945.

GEER, J. H. Phobia treated by reciprocal inhibition. *Journal of Abnormal and Social Psychology,* 1964, 69, 642–645.

GROSSBERG, J. M. Behavior therapy: A review. *Psychological Bulletin,* 1964, 62, 73–88.

HATHAWAY, S. R. A coding system for MMPI profile classification. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

JACOBSON, E. *Progressive relaxation.* (2nd ed.) Chicago: Univer. Chicago Press, 1938.

MCGRAW, R. B., & OLIVEN, J. F. Miscellaneous therapies. In S. Arieti (Ed.), *American handbook of psychiatry.* New York: Basic Books, 1959.

WOLBERG, L. R. *The technique of psychotherapy.* New York: Grune & Stratton, 1954.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford Univer. Press, 1958.

WOLPE, J., SALTER, A., & REYNA, L. J. *The conditioning therapies.* New York: Holt, Rinehart & Winston, 1964.

(Received April 2, 1965)

---

[2] Residual tension as used by Jacobson refers to fine, tonic muscular contractions.

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 165–168

INSTRUMENTAL LEARNING WITH SOCIOPATHS:

A TEST OF CLINICAL THEORY

ROY W. PERSONS AND JAMES L. BRUNING[1]
Ohio State University *Ohio University*

27 incarcerated sociopaths, 27 incarcerated nonsociopaths, and 27 normals were instructed to draw 3-in lines while blindfolded. Knowledge of results was the verbal response "correct," "too long," or "too short," plus mild electric shock for incorrect responses. All groups showed rapid improvement during the treatment trials. Contrary to some recent experimental findings, the incarcerated sociopaths improved with reinforcement and, in addition, demonstrated the most rapid acquisition and greatest resistance to extinction.

Cleckley's (1950) theory of psychopathy, which emphasizes the psychopath's inability to profit from experience, has recently received considerable experimental support. Lykken (1957), in an avoidance-conditioning study, concluded that sociopaths were defective in their ability to learn an avoidance response. Similarly, Johns and Quay (1962) found that psychopaths were less sensitive to secondary reinforcement than neurotics, and suggested that their results, together with Lykken's, "might be taken as evidence for a general factor of poor conditionability in the psychopath, irrespective of the type of conditioning." A replication of the Johns and Quay (1962) study by Quay and Hunt (1965) arrived at essentially the same conclusion.

However, as pointed out by Persons and Persons (1965), the studies by Johns and Quay, and Lykken have numerous methodological problems (e.g., statistical and design procedures, group selection, and initial ceiling effect), and thus should not be evaluated at face value. The present investigation is also designed to test the sociopath's amenability to conditioning but with the aim of avoiding the difficulties present in the previous studies.

METHOD

Subjects

A total of 95 male subjects were selected, 68 from volunteers at a state reformatory, Boys' Industrial School, and 27 from Ohio University.[2] There were 27 incarcerated sociopaths (IS), 27 incarcerated nonsociopaths (INS), 27 college freshmen (Normals), and a control group of 14 incarcerated subjects.

All 27 sociopaths carried a psychiatric diagnosis of sociopathic personality, which had been arrived at through interviewing, review of case history, administration of the Rorschach, and the Minnesota Multiphasic Personality Inventory. In addition, all 27 of the sociopaths finally selected for participation were below the institutional mean on neuroticism and above the mean on psychopathy on the Delinquency scale (Peterson, Quay, & Tiffany, 1961). The sociopaths were also below the institutional mean on the Taylor Manifest Anxiety (*MA*) scale (Taylor, 1953). The 27 INS were selected from volunteers carrying a psychiatric diagnosis other than sociopathic personality. There were no psychotics in this group. The control group of 14 incarcerated subjects were selected at random from the institutional files without respect to psychiatric diagnosis. The 27 Normals were drawn from a freshman psychology class at Ohio University. The mean age of the incarcerated subjects was approximately 17, and the mean age of the university subjects was approximately 18.

Apparatus

A detailed description of the apparatus may be found in Becker, Mussina, and Persons (1963). The essential part of the apparatus was a large sheet of Plexiglas with a horizontal groove for drawing lines. An inkless pen, standard electric clock, Superior Powerstat with wristband electrodes, blindfold, and stopwatch completed the apparatus.

Procedure

The procedure for the study was the same as that employed by Becker et al. (1963). The subjects were run individually and seated before a table in a quiet room. The experimenter read the instructions (Greenspoon & Foreman, 1956) aloud to the subject, who was then blindfolded, and the apparatus was uncovered. The wristband electrodes were fastened to the

---

[1] The authors wish to thank David Mettee, Roger Uhl, Richard White, and the officials of Boys' Industrial School for their kind assistance in this project.

[2] There were 987 boys incarcerated at the reformatory. The volunteers were selected by the institutional psychologist from a group of 411 boys for whom there were psychological and background data available.

TABLE 1

TREATMENT CONDITIONS FOR EXPERIMENTAL GROUPS

Group	Trials		
	1–10	11–35	36–60
IKR-5	No KR	Immediate KR	No KR
IKR-25	No KR	Immediate KR	No KR
DKR	No KR	Delayed KR	No KR

subject's left wrist and he was instructed to report when he felt a "tickling" sensation. When threshold was reported, the experimenter turned off the shock, increased the level by two additional volts, and returned the left arm to the subject's lap. The subject's task was to draw a 3-inch line while blindfolded. The pen was placed in the subject's preferred hand and guided to the start point. The subject was given a ready signal and the first trial began. Each subject drew 60 lines during 60 consecutive trials. Responses were recorded to the nearest ¼ inch; any response from 2¾ inches to 3¼ inches was considered correct.

The IS, INS, and Normal subjects were subdivided into three equal groups (see Table 1). Thus, each group was composed of 9 IS, 9 INS, and 9 Normal subjects. The knowledge of results (KR) administered for incorrect responses consisted of an appropriate verbal statement by experimenter ("too short" or "too long") with a simultaneous shock of 1-second duration. KR for correct responses was the verbal statement, "correct." The draw signal was the click of the reset switch on the standard electric clock. Response completion was defined by the action of the subject's resting his arm on the table.

To insure that all subjects began performing at the same initial level, no KR was given to any of the subjects during the first 10 trials. The following procedures were used for the various groups: (*a*) IKR-5: 5 seconds following response completion, the experimenter moved the subject's hand back to the start point. Draw signals were given every 15 seconds. During Trials 11–35, KR was given immediately upon response completion. (*b*) IKR-25: 25 seconds following response completion, the experimenter moved the subject's hand back to the start point. During Trials 11–35, KR was given immediately upon response completion. (*c*) DKR: 25 seconds following response completion, the experimenter moved the subject's hand back to the start point. Draw signals were given every 35 seconds. During Trials 11–35, KR was given 20 seconds after response completion.

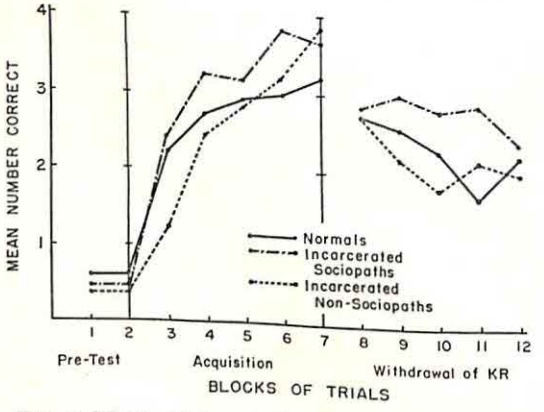


FIG. 1. Mean total correct responses over blocks of five trials. KR was not administered during Blocks 1–2 and 8–12. All subjects received KR during Blocks 3–7.

RESULTS

The performance measures of mean total correct responses over blocks of five trials for all groups are plotted in Figure 1. The initial analysis of variance, computed for the first 10 trials when no KR was given to any of the subjects, showed all of the F values to be less than 1. From this it can be concluded that the groups were performing at essentially the same level before entering the treatment trials.

To insure that performance did not improve as a function of trials rather than KR, 14 subjects were given the entire 60-trial sequence under conditions of no KR. Since this group did not show any increment in performance during the treatment trials, their results were not included in the data analysis or in Figure 1.

The treatment conditions resulted in rapid performance improvement for all three experimental groups. The analysis of variance presented in Table 2 indicates that the effects of Psychological Classification (N/INS/IS) and also the Trials × Psychological Classifi-

TABLE 2

ANALYSIS OF VARIANCE FOR NUMBER OF CORRECT RESPONSES DURING THE TREATMENT TRIALS

Source	*df*	*MS*	*F*
Total	404		
Between subjects	80		
KR	2	.99	
N/INS/IS	2	9.01	5.72*
KR × N/INS/IS	4	1.36	
Error (b)	72	1.58	
Within subjects	324		
Trials	4	31.15	24.24**
T × KR	8	.81	
T × N/INS/IS	8	3.26	2.54*
T × KR × N/INS/IS	16	1.84	
Error (w)	288	1.28	

* $p < .01$.
** $p < .001$.

TABLE 3

ANALYSIS FOR NUMBER OF CORRECT RESPONSES DURING EXTINCTION

Source	*df*	*MS*	*F*
Total	404		
Between subjects	80		
KR	2	6.53	
N/INS/IS	2	13.65	3.70**
KR × N/INS/IS	4	1.48	
Error (b)	72	3.70	
Within subjects	324		
Trials	4	4.30	2.27*
T × KR	8	1.90	
T × N/INS/IS	8	3.78	2.00**
T × KR × N/INS/IS	16	1.18	
Error (w)	288	1.89	

* $p < .10$.
** $p < .05$.

cation were significant. Inspection of Figure 1 indicates that the IS group demonstrated the most rapid acquisition while the INS was slowest. As in previous studies (Bilodeau & Bilodeau, 1958; Bilodeau & Ryan, 1960; Becker et al., 1963), delay of KR and intertrial interval were not significant variables.

Following the 25 nonreinforced trials, the performance of none of the groups had declined to the pretreatment level. As can be seen from the analysis of variance presented in Table 3, there was a significant difference between the three psychological classification groups and also a significant Trials × Psychological Classification interaction.

DISCUSSION

It should be clear from the evidence that sociopaths in this study were not defective in conditionability. In fact, the sociopathic group conditioned faster and showed less performance decrement during extinction. Thus, the results of this experiment are at odds with Johns and Quay's (1962) statement that psychopaths condition poorly regardless of the type of conditioning. Johns and Quay used verbal praise for reinforcement, while Lykken (1957) used anxiety reduction from avoiding a strong electric shock as reinforcement. The reinforcement conditions in the present experiment were verbal knowledge of results plus electric shock.

It is not surprising that the prisoners volunteered to participate in the study when the rather routine way of institutional life is considered. A large number of these inmates had neither a work nor a school assignment, and they sought out almost any activity that would occupy their time. The inmate subjects were as fully cooperative as the university students who were given course points for their participation.

An unexpected finding of the study was that the INS group initially conditioned much slower than the IS group, but in later trials equaled the performance of the sociopaths. A plausible explanation for this difference in rate of conditioning is that initial anxiety of the INS group impeded performance, but as anxiety declined performance improved concomitantly (Lazarus, Deese, & Osler, 1952). A reexamination of the institutional records of the subjects in the INS group showed that approximately 67% of these subjects had been diagnosed as neurotics with anxiety reactions, which lends some support to the hypothesis. In addition, these subject's Taylor *MA* scale scores were found to be significantly above the mean score of the other inmate subjects.

The typical statement that sociopaths are incapable of profiting from experience or that they are not amenable to conditioning is given as supporting evidence for their receiving no psychological or psychiatric treatment (Pennington, 1954). Bender (1947) asserts that treatment for sociopaths is doomed to failure. However, more recently, favorable results in treating sociopaths have been reported (Persons, 1965; Sturup, 1952). It has been clearly demonstrated that sociopaths are susceptible to conditioning, and that in some situations they can, and do, profit from prior experience. Perhaps a fruitful line for future investigation should consider not only different experimental and psychotherapy situations, but also the factors which operate to make reinforcements effective or ineffective.

REFERENCES

BECKER, P. W., MUSSINA, C. W., & PERSONS, R. W. Intertrial interval, delay of knowledge of results, and motor performance. *Perceptual & Motor Skills*, 1963, 17, 559–563.

BENDER, L. Psychopathic behavior disorders in children. In R. M. Lindner & R. V. Seliger (Eds.), *Handbook of correctional psychology*. New York: Philosophical Library, 1947.

BILODEAU, E. A., & BILODEAU, I. M. Variation of temporal intervals among critical events in five studies of knowledge of results. *Journal of Experimental Psychology*, 1958, **55**, 603–612.

BILODEAU, E. A., & RYAN, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *Journal of Experimental Psychology*, 1960, **59**, 414–419.

CLECKLEY, H. *The mask of sanity*. St. Louis: Mosby, 1950.

GREENSPOON, J., & FOREMAN, S. Effect of delay of knowledge of results on learning a motor task. *Journal of Experimental Psychology*, 1956, **51**, 226–228.

JOHNS, J. H., & QUAY, H. C. The effect of social reward on verbal conditioning in psychopathic and neurotic military offenders. *Journal of Consulting Psychology*, 1962, **26**, 217–220.

LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychological Bulletin*, 1952, **49**, 293–317.

LYKKEN, D. F. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, 1957, **55**, 6–10.

PENNINGTON, L. A. Psychopathic and criminal behavior. In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology*. New York: Ronald, 1954.

PERSONS, R. W. Psychotherapy with sociopathic offenders: An empirical evaluation. *Journal of Clinical Psychology*, 1965, **21**, 204–207.

PERSONS, R. W., & PERSONS, C. E. Some experimental support for psychopathic theory: A critique. *Psychological Reports*, 1965, **16**, 745–749.

PETERSON, D. R., QUAY, H. C., & TIFFANY, T. L. Personality factors related to juvenile delinquency. *Child Development*, 1961, **32**, 355–372.

QUAY, H. C., & HUNT, W. A. Psychopathy, neuroticism and verbal conditioning: A replication and extension. *Journal of Consulting Psychology*, 1965, **29**, 283.

STURUP, G. K. The treatment of criminal psychopaths in Hertedvester. *British Journal of Medical Psychology*, 1952, **25**, 31–38.

TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1953, **48**, 285–290.

(Received April 14, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 169–180

EFFECTS ON RECALL OF HYPNOTIC INDUCTION, MOTIVATIONAL SUGGESTIONS, AND SUGGESTED REGRESSION:

A METHODOLOGICAL AND EXPERIMENTAL ANALYSIS[1]

THEODORE XENOPHON BARBER AND DAVID SMITH CALVERLEY

Medfield Foundation, Harding, Massachusetts

90 Ss were assessed for recall of syllables (learned 2 mo. previously) under 1 of 9 experimental conditions in a 3 × 3 factorial design. The 3 levels of the 1st independent variable were: hypnotic induction, eyes closed; no induction, eyes closed; and no induction, eyes open. The 3 levels of the 2nd independent variable were: recall without suggestions; recall with motivational suggestions; and recall with suggestions to regress to the time of original learning. Recall was not affected by the independent variables or their interaction. A comparable number of Ss in each of the 9 experimental groups testified post-experimentally that, during the experiment, they had imagined, felt, and believed they were back in the original learning situation.

A series of experiments has ostensibly demonstrated that recall of remote memories or of previously learned material can be facilitated by first administering a hypnotic induction procedure and then administering either (*a*) motivational suggestions for heightened recall or (*b*) suggestions to regress to the time of original learning. If these findings are valid, hypnotic procedures which include motivational suggestions or regression suggestions should prove useful in various situations, such as in psychotherapy, in producing heightened recall of important life-history material. However, the findings cannot as yet be accepted as defintive; all pertinent experiments conducted up to this time are open to one or more serious methodological objections. The present paper first specifies methodological problems relevant to research in this area and then presents an experiment that was designed to avoid inadequacies in method.

METHODOLOGICAL CONSIDERATIONS

In this section are discussed, in consecutive order, methodological problems pertaining to assessment of the effects on recall of (*a*) motivational suggestions given under hypnotic conditions and of (*b*) hypnotic age-regression suggestions. Following this, consideration is given general problems in method pertaining to (*c*) control of experimenter bias and (*d*) control of bias in selection of subjects.

Motivational suggestions for heightened recall. Stalnaker and Riddle (1932) and Hammer (1954) found that suggestions of high motivation and increased ability given subsequent to a hypnotic induction were effective, as compared to a nonsuggestion waking condition, in facilitating recall of previously learned material. However, the possibility was not considered that subjects might have performed as proficiently under the waking condition as under the hypnotic condition if motivational suggestions for enhanced recall had been given under both conditions. Since the investigations confounded two independent variables, namely, (*a*) hypnotic induction (or the "hypnotic trance state") and (*b*) motivational suggestions intended to elicit a high level of performance, no conclusions are possible concerning the effects of hypnotic induction alone, or of motivational suggestions alone, or of the interaction of hypnotic induction with motivational suggestions.

White, Fox, and Harris (1940) reported that suggestions for enhanced recall were more effective when given under hypnotic, rather than under waking conditions, in improving recall of poetic material, but not of nonsense syllables or of motion picture scenes. In a series of studies, Rosenthal (1944) found that motivational suggestions for improved recall of meaningful and nonsense material at times were more effective and at times were not more effective when given under hypnotic, rather than under waking conditions.

In the White et al. and Rosenthal studies the subjects were apparently tested for recall with their eyes open under the waking treatment and with their eyes closed under the hypnotic treatment.

---

[1] This research was supported by a grant (MH-07003) from the National Institute of Mental Health, United States Public Health Service. We are indebted to Frederick J. Ryan of Worcester Junior College for assistance in the conduct of the experiment.

Since recall may be more proficient when the eyes are closed, the possibility was not excluded that the heightened recall that was at times obtained under the hypnotic treatment was due, not to hypnotic induction (or "hypnotic trance"), but to the removal of distractions which occurs when the eyes are closed (cf. Barber, 1965a).

It follows from the above considerations that further research is needed to assess the effects on recall of the following variables independently: hypnotic induction; motivational suggestions for heightened recall; and recall with eyes closed. In the present experiment the effects of these variables were evaluated separately and in all possible combinations.

Age-regression suggestions. Several studies (e.g., Ås, 1962; True, 1949), reviewed in detail elsewhere (Barber, 1962b), indicate that recall can be facilitated by administering a hypnotic induction and then suggesting to the subject that he regress to the time of original learning. However, each of these studies lacked an important control; in no instance were suggestions to regress given to awake subjects. Since the regression suggestions were always confounded with the hypnotic induction, no conclusions can be drawn concerning the effects on recall of hypnotic induction alone, or of regression suggestions alone, or of the interaction of hypnotic induction with regression suggestions. Clearly, further research is needed to assess the effects on recall of regression suggestions given both with and without a preceding hypnotic induction. The present experiment was designed to meet this need.

Control of experimenter bias. There is evidence to indicate that experimenters find it difficult to give identically worded suggestions to hypnotic and waking subjects in the same tone of voice, and that the biased tone of voice in which the suggestions are presented effects the subjects' responses (Barber & Calverley, 1964b; Troffer & Tart, 1964). To control this factor in our experiment, the suggestions were administered in exactly the same way to hypnotic and waking subjects by means of a tape recording of the experimenter's voice (Barber & Calverley, 1964a).

Control of bias in selection of subjects. In some of the investigations in this area, subjects allocated to the hypnotic treatment, but not to the waking control treatment, were preselected for "hypnotizability." This selection procedure is open to two serious objections, as follows:

1. To certify that subjects allocated to the hypnotic treatment met criteria for "hypnotizability," preliminary assessment or "training" sessions were conducted with them. Discrepancies in performance during the critical session may thus have been contingent, not upon the experimental group being in "hypnosis" and the control group being in the "awake state," but upon the experimental group having participated in the preliminary sessions and having acquired useful experience. For instance, in the preliminary sessions, the experimental group may have formed a friendly relationship with the experimenter and may have become familiar with the experimental setting. These, and possibly other factors carried over from the assessment or "training" sessions to the experiment proper, could have the general effect of making the experimental subjects more at ease or less anxious than the controls and of enhancing their motivation and capability to give a maximal performance on the tests for recall.

2. The explicit criterion for selection of the experimental group—that the subjects were "hypnotizable"—appears difficult to differentiate from the interrelated implicit criterion that the subjects were highly responsive to suggestions with or without "hypnotic induction." By thus confounding the experimental conditions with preexisting differences among subjects with respect to "hypnotizability" or "suggestibility," it could not be ascertained whether the experimental subjects were more responsive than the controls to suggestions for heightened recall because they received the suggestions when they were in "hypnotic trance," or because they were more responsive to suggestions with or without "hypnosis."

In other experiments in this area, selected "good" hypnotic subjects were used as "their own controls," that is, were tested under both the hypnotic and waking conditions. This procedure is open to the following criticism: there is evidence to indicate that some "good" hypnotic subjects surmise that the experimenter expects or wants better results under the hypnotic condition and, desiring to please the experimenter, purposely give their best possible performance under the hypnotic condition and purposely give an inferior performance under the waking condition (Barber, 1962a; Pattie, 1935; Schrenck-Notzing, 1896; Sutcliffe, 1960; Wolberg, 1948, p. 49).

The above considerations suggest that further studies should allocate different subjects to the hypnotic and waking treatments and should avoid confounding the hypnotic treatment with such variables as prior "training" and preexisting high suggestibility. In accordance with these considerations, in the present experiment, unselected and untrained subjects from a homogeneous population were assigned *at random* to the treatments.

METHOD

Experimental Design

All subjects first learned a list of syllables. Two months later they were tested individually for recall of the syllables under one of nine experimental conditions. The nine experimental conditions constituted the cells of a 3 × 3 factorial as illustrated in Table 1. The first independent variable (Experimental Treatments) was at three levels: hypnotic induction, eyes closed; no hypnotic induction, eyes closed; and no hypnotic induction, eyes open. The second independent variable (Conditions of Recall) was also at three levels: recall without suggestions; recall with motivational suggestions; and recall with regression suggestions.

Subjects

Subjects were 90 freshmen and sophomores (77 males and 13 females) at Worcester Junior College who were enrolled in the introductory psychology course taught by Frederick J. Ryan. The 90 students were randomly assigned to the nine experimental groups with 10 to each group.

Original Learning

When assembled in classes during the first week of October 1964, the students were given 12 syllables to learn. The syllables were introduced as follows:

The instructor in the class (F. J. Ryan) had been discussing the history of psychology and during this particular class session lectured on Ebbinghaus and the learning of nonsense material. At the conclusion of the lecture the students were told that, to better understand what was involved in memorizing nonsense syllables, they should try to learn a list themselves. A 22 × 8-inch white cardboard containing 12 syllables was then affixed on the front wall. The students were told they would have 5 minutes to learn the list. The list was removed after 5 minutes and the students were asked to write the syllables on a sheet of paper and to sign their name. After 3 minutes elapsed, the papers were collected. On the average, the students wrote 10 syllables of which 9 were the same as those on the original list.[2]

The 12 syllables were selected in accordance with the following considerations. It was desired that the material should be standardized, should consist of equal units for convenience of scoring, and should be of equal difficulty for all subjects. Standardized lists of nonsense syllables as compiled by Glaze (1928) best meet these criteria. However, since it was also desired that the material should be as meaningful as possible, the 12 syllables were randomly selected from among those that Glaze had found to be of 93% association value (Hilgard, 1951, p. 541). These syllables were: KAF, PUR, REM, PAV, LEC, NOM, LOY, YAC, FIV, TIR, MUR, and LUN.

Introduction to Experimental Sessions

Two months after the original learning, the students were informed by the instructor that in order to fulfill the course requirements, they would participate in an experiment. They were further told that: by participating in the experiment they would increase their understanding of psychology and what it is like to be a subject in a psychological investigation; the experiment would begin the following day; it would be conducted by a former instructor at the college (D. S. Calverley); some subjects would be hypnotized and others would serve in a control group and would not be hypnotized; the experiment would be explained to them after it was completed; and, during the 10-day period the experiment was in progress, they were not to discuss it with each other. The instructor concluded by carefully explaining to the class why the experiment would be "ruined" if the students talked about it. (At the end of each subject's test-session, the subject was again admonished, this time by the experimenter, not to "ruin the experiment" by discussing it with the other students. Each subject stated that he understood the importance of this admonition and promised to adhere to it.)

Experimental Treatments

On each of the 10 experimental days, one subject was tested from each experimental group. Each subject was tested individually by one experimenter (DSC) under one of three Experimental Treatments, and also under one of three Conditions of Recall, as illustrated in Table 1. The three experimental treatments were as follows:

Hypnotic induction, eyes closed. Subjects allocated to Groups 1, 2, and 3 were first told:

"You have been assigned to the hypnosis group and you will be deeply hypnotized. The instructions for this experiment are on this recording. Listen carefully to the recording and follow the instructions."

A standardized 10-minute hypnotic induction procedure was then administered by a tape recording of the experimenter's voice. This induction procedure, which was patterned closely after the hypnotic procedures of Marcuse (1959, pp. 52–53) and Weitzenhoffer and Hilgard (1959, pp. 13–18), included instructions designed to produce positive rapport and positive motivation to perform well in the hypnotic situation, suggestions of eye-heaviness and eye-closure, repeated suggestions of relaxation, drowsiness, and sleep, suggestions that the subject was entering a deep hypnotic trance, and suggestions that he would now be able to respond easily to further suggestions.[3]

Upon completion of the hypnotic induction procedure, Groups 1, 2, and 3 were told "Keep your eyes closed" and were then tested immediately for recall of the previously learned syllables under one of three Conditions of Recall, as described subsequently.

No induction, eyes closed. Subjects allocated to Groups 4, 5, and 6 were told to keep their eyes closed during the experiment but were not exposed to a hypnotic induction. Instructions for these groups were:

---

[2] The reader should note that the students were *not* told that they would again be asked to recall the nonsense syllables at a later time. On the contrary, the impression was given that, after they had turned in their list at the end of the class session, the nonsense syllables were a closed matter.

[3] A verbatim account of the hypnotic induction procedure may be obtained without charge by writing to the authors or for a fee from the American Documentation Institute. Order Document No. 8320 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. Remit in advance $1.25 for photocopies or 35-mm. microfilm and make checks payable to: Chief, Photoduplication Service, Library of Congress.

> You have been assigned to the waking control group and you will not be hypnotized. The instructions for this experiment are on this recording. I want you to close your eyes and to keep them closed during the experiment. Now close your eyes and keep them closed and listen carefully to the recording and follow the instructions.

Groups 4, 5, and 6 were then tested immediately for recall of the previously learned syllables under one of three Conditions of Recall.

No induction, eyes open. Subjects assigned to Groups 7, 8, and 9 were not exposed to a hypnotic induction and were not asked to close their eyes. Instructions were:

> You have been assigned to the waking control group and you will not be hypnotized. The instructions for this experiment are on this recording. Listen carefully to the recording and follow the instructions.

Groups 7, 8, and 9 were then tested immediately under one of three Conditions of Recall, as described next.

Conditions of Recall

Recall without suggestions. Group 1 (Hypnotic induction, eyes closed), Group 4 (No induction, eyes closed), and Group 7 (No induction, eyes open) were asked to recall the previously learned material without receiving motivational suggestions or suggestions to regress. The instructions, which were administered by a tape recording, were:

> Now I want you to listen very carefully to what I say. At the beginning of the semester, during the first week in October, Mr. Ryan showed you a large card with nonsense syllables on it and he asked you to learn the nonsense syllables. Now, during the next 3 minutes, I want you to recall as many of the nonsense syllables as you can. Spell out the nonsense syllables aloud, one by one. Start now.

The subjects were given 3 minutes for recall.

Recall with motivational suggestions. Subjects in Group 2 (Hypnotic induction, eyes closed), Group 5 (No induction, eyes closed), and Group 8 (No induction, eyes open) were given 3 minutes to recall the nonsense syllables after receiving the following motivational suggestions for heightened recall by means of a tape recording:

> Now I want you to listen very carefully to what I say. At the beginning of the semester, during the first week in October, Mr. Ryan showed you a large card with nonsense syllables on it and he asked you to learn the nonsense syllables. If you really try very hard to remember the nonsense syllables, you will be able to recall them. Even though you may feel that they are difficult to remember, you'll be surprised how many of them you will be able to recall if you really try hard to recall as many of the nonsense syllables as you can. Try very hard to recall them. Don't assume that you can't remember them. Other students were able to do surprisingly well on this test. You too can remember them very well if you really try. Now, during the next 3 minutes, I want you to recall as many of the nonsense syllables as you can. Spell out the nonsense syllables aloud, one by one. Start now.

Recall with regression suggestions. Subjects in Group 3 (Hypnotic induction, eyes closed), Group 6 (No induction, eyes closed), and Group 9 (No induction, eyes open) were given suggestions to regress to the time when they originally learned the nonsense syllables (2 months previously). The suggestions, given by means of a tape recording, were as follows:

> Now I want you to listen very carefully to what I say. I want you to imagine vividly that you are going back in time to the beginning of the semester, during the first week in October, when Mr. Ryan showed you a large card with nonsense syllables on it and he asked you to learn the nonsense syllables. Time is now going backwards. The days are going backward. It is now November 10, 1964. The days are going backward. It is now November 1, 1964. You are going back in time. It is now October 20th. Time is going back. It is now the second week in October. You are going back in time. It is now the first week in October. It is the first week in October and you are in Mr. Ryan's class. Mr. Ryan has given you a list of nonsense syllables to learn. It is the first week in October, you are in Mr. Ryan's class, and you have just learned the nonsense syllables. Now answer these questions aloud. Answer these questions aloud, clearly, and distinctly.
>
> In whose class are you? (10-second pause)
>
> What is the month and the week? (10-second pause)
>
> What have you just learned? (10-second pause)
>
> Now spell out the nonsense syllables that you have just learned. During the next 3 minutes spell out aloud each of the nonsense syllables that you have learned. Spell them out aloud one by one. Start now.

Immediately following the 3-minute recall period, Groups 3, 6, and 9 were told: "Now you are back to the present. You are now back to the present."

Assessment of Hypnotic Subjects' Response to Barber Suggestibility Scale

Upon completing the above, subjects in Groups 4–9 (No induction) were told that the experiment was over. Subjects in Groups 1–3 (Hypnotic induction) were given additional suggestions of relaxation, drowsiness, and sleep for 2 minutes, then assessed on response to the Barber Suggestibility Scale (Barber, 1965c; Barber & Calverley, 1963; Barber, Karacan, & Calverley, 1964) and, finally, told that the experiment was over.

Postexperimental Subjective Reports (Imagined, Felt, and Believed)

Immediately after it was stated that the experiment was over, all 90 subjects were given a 3-item dittoed questionnaire to answer, which was worded as follows:

> Please answer the following questions by placing a check mark (V) next to the appropriate answer.
>
> 1. During the experiment, when you were asked to recall the nonsense syllables, to what extent did you *imagine* that it was the first week in October, that you were in Mr. Ryan's class, and that you had just learned the nonsense syllables:
>
> [A] _____I thoroughly imagined this.
> [B] _____I imagined this to a great extent.
> [C] _____I imagined this to some extent.
> [D] _____I did not imagine this.
>
> Item 2 was identical with Item 1 with the exception that the word *feel* (or *felt*) was substituted for the word *imagine* (or *imagined*).
>
> Item 3 was also identical with Item 1 with the exception that the word *believed* was substituted for the word *imagined*.
>
> Each of the three items (*imagined, felt,* and *believed*) was scored as follows: Weights of 3, 2, 1, and 0 were assigned to A, B, C, and D answers, respectively.

RESULTS AND DISCUSSION

First the authors shall present the results pertaining to the primary question posed in this experiment, namely, what were the effects on recall of the two major independent variables (Experimental Treatments and Conditions of Recall)? Next, we shall turn to three supplementary questions: (*a*) Were suggestions to regress effective in producing regression as indicated by subjects' answers to questions concerning where they were in place and time and what they were doing? (*b*) What were the effects of the independent variables on postexperimental subjective reports? (*c*) With respect to Groups 1–3, did the better hypnotic subjects (those obtaining higher scores on the Barber Suggestibility Scale) differ from the poorer hypnotic subjects in recall, in degree of regression, or in postexperimental subjective reports?

Effects of Independent Variables on Recall

Each of the nonsense syllables proffered by the subject during the experimental session was scored as correct if it was the same as one of the nonsense syllables that the subject had written down during the original learning session 2 months previously. The mean recall scores (average number of correct nonsense syllables) for each experimental group are presented in Table 1 together with an analysis of variance of the scores. This table shows that all experimental groups correctly recalled an average of *less than one* nonsense syllable. The number of syllables correctly recalled did not differ significantly among (*a*) subjects who had received a hypnotic induction and whose eyes were closed, (*b*) subjects who had not received an induction and whose eyes were closed, and (*c*) sub-

TABLE 1

MEAN NUMBER OF NONSENSE SYLLABLES CORRECTLY RECALLED (RECALL SCORES) AND ANALYSIS OF VARIANCE OF RECALL SCORES

	Experimental treatments			Average
	Hypnotic induction, eyes closed	No induction, eyes closed	No induction, eyes open	
Conditions of recall				
Recall without suggestions	.4 (Group 1)	.9 (Group 4)	.1 (Group 7)	.5
Recall with motivational suggestions	.2 (Group 2)	.1 (Group 5)	.7 (Group 8)	.3
Recall with regression suggestions	.5 (Group 3)	.4 (Group 6)	.2 (Group 9)	.4
Average	.4	.5	.3	

Source of variance	*df*	*MS*	*F*
Experimental treatments (A)	2	.14	
Conditions of recall (B)	2	.14	
A × B	4	1.30	2.24
Error	81	.58	
Total	89		

jects who had not received an induction and whose eyes were open (nonsignificant main effect for Experimental Treatments). Furthermore, the recall scores did not differ significantly among (*a*) subjects who recalled without suggestions, (*b*) subjects who recalled with motivational suggestions, and (*c*) subjects who recalled with regression suggestions (nonsignificant main effect for Conditions of Recall). The interaction of the independent variables (Experimental Treatments with Conditions of Recall) was also nonsignificant. These nonsignificant effects reflected the following: The overwhelming majority of subjects (88%) correctly recalled either none or one of the nonsense syllables; the 11 subjects (12%) who recalled either two or three of the syllables were distributed more or less evenly over the nine experimental groups. (No subject correctly recalled more than three syllables.)[4]

These results are in harmony with recent experiments presented by Lyon-James (1957) and by Leonard (1963). In the Lyon-James experiment, 12 selected, good hypnotic subjects were assessed for recall of meaningful material under hypnotic conditions, and an additional 12 selected, good hypnotic subjects were assessed for recall under waking conditions, with motivational suggestions for high recall given under both conditions. The tests for recall were given immediately after the subjects had learned the material, after 3 weeks, and again after 9 weeks. The hypnotic and waking groups did not differ significantly in either immediate or delayed recall of the material. In the Leonard experiment, five highly selected, good hypnotic subjects who were regressed to the time when they had originally learned a list of paired associates (2 weeks previously) did not relearn the associates more proficiently than 10 unselected nonhypnotic subjects who were not given suggestions to regress to the time of original learning.

The results obtained in our experiment, and in the recent experiments of Lyon-James and of Leonard, do not exclude the possibility that hypnotic induction, or eyes closed, or motivational suggestions, or regression suggestions, or some combinations of these variables may at times facilitate recall. As noted in the introduction to the present paper, several studies (Ås, 1962; Rosenthal, 1944; White et al., 1940) found that recall was at times facilitated by motivational suggestions or regression suggestions given subsequent to a hypnotic induction procedure. The latter studies differed from each other, and also from our study and from the studies of Lyon-James and of Leonard, in the methods used to select subjects, in the wording of instructions and suggestions, and in many other methodological aspects. Clearly, further research is needed to determine which of the many procedural differences among these investigations were important and which were extraneous in producing both the positive and the negative results. Future studies, how-

---

[4] A series of additional statistical analyses were also performed on the Recall scores. These included the following:

1. Two analyses of covariance of the scores presented in Table 1. In the first covariance analysis, the covariate was the total number of syllables that the subject had written down during the original learning session 2 months previously. In the second covariance analysis, the covariate was the number of syllables that the subject had recalled correctly during the original learning session.

2. An analysis of variance and two analyses of covariance of the number of nonsense syllables given by the subjects during the experimental session which were either correct or two-thirds correct. In these analyses, a syllable received a score of 1 if at least 2 of the 3 letters were identical with and in the same positions as the letters in one of the syllables recalled by the subject during the original learning session. (In the covariance analyses, the covariates were the same as in Paragraph 1 of this footnote.)

These analyses yielded the same results as the analysis presented in the body of this paper. That is, the main effects and the interaction of the independent variables did not significantly affect recall of the previously learned material. These nonsignificant effects reflected the fact that, irrespective of the scoring method that was used, the overwhelming majority of subjects gave no more than two correct syllables and the few subjects that scored above 2 were distributed more or less evenly over the nine experimental groups.

To ascertain whether the independent variables affected the degree to which the subjects *tried* to give nonsense syllables (Attempt scores), an additional analysis of variance was performed on the total number of nonsense syllables elicited from the subjects during the 3-minute experimental recall period, irrespective of whether the syllables were correct or incorrect. This analysis also yielded nonsignificant *F* ratios.

TABLE 2

NUMBER (AND PERCENTAGE) OF SUBJECTS PASSING REGRESSION QUESTIONS

	Question			
	1. In whose class are you?	2. What is the month and the week?	3. What have you just learned?	1–3
Group 3 (Hypnotic induction, eyes closed)	10 (100%)	10 (100%)	10 (100%)	10 (100%)
Group 6 (No induction, eyes closed)	9 (90%)	6 (60%)	10 (100%)	6 (60%)
Group 9 (No induction, eyes open)	9 (90%)	9 (90%)	8 (80%)	8 (80%)

ever, should take into consideration the experimental controls discussed in the methodological introduction to the present paper. For instance, since there is evidence to indicate that experimenters may bias the outcome by varying the tone of voice in which they administer suggestions to "hypnotized" and "awake" subjects (Barber & Calverley, 1964b; Troffer & Tart, 1964), further research should consider controlling experimenter bias in the same way as in the present experiment, namely, by using a tape recording to administer instructions and suggestions.

Effectiveness of Suggestions to Regress

Group 3 (Hypnotic induction, eyes closed), Group 6 (No induction, eyes closed), and Group 9 (No induction, eyes open) were given suggestions to regress to the time (2 months previously) when they had originally learned the nonsense syllables. Following the suggestions to regress, the subjects in these groups were asked three questions: (*a*) In whose class are you?; (*b*) What is the month and the week?; and (*c*) What have you just learned? These questions provided three interrelated criteria for regression, as follows: In reply to the first, second, and third questions, respectively, the subject who is regressed states that (*a*) he is in Mr. Ryan's class, (*b*) it is the first week in October, and (*c*) he has just learned a list of nonsense syllables. Were the suggestions to regress effective in producing regression as indicated by these criteria? Table 2 presents the pertinent data. The final column in this table shows that 100%, 60%, and 80% of the subjects in Groups 3, 6, and 9, respectively, answered each of the three questions as if they were regressed. A chi-square test showed that the differences between the groups in the number of subjects passing each of the three criteria for regression fell short of significance at the 5% level of confidence. The import of these experimental results is as follows:

1. Sixty percent and 80% of the subjects in the two groups that did not receive a hypnotic induction met each of the three criteria for regression (Table 2, final column). This outcome, indicating that suggestions to regress are effective in producing regression in the majority of "waking" subjects, does not contradict any previous study known to us. Although in several previous studies nonhypnotic subjects were asked to simulate or to pretend they were regressed, in no prior investigations were nonhypnotic subjects given the same suggestions to regress as were given to the hypnotic subjects (see Barber, 1961, 1962b, for a review of the literature).

2. It needs to be emphasized that all prior experiments failed to give regression suggestions to nonhypnotic subjects and, further, that all prior experiments included, among the criteria for regression, statements from the hypnotic subjects that they were in another place and at a previous time. Clearly, no conclusions concerning the effectiveness of hypnotic induction (or "hypnosis") in producing acceptance of suggestions to regress can be drawn from any previous experiment until it is replicated with the addition of a nonhypnotic group that is given identical suggestions to regress and is assessed on the same criteria for regression as the hypnotic group.

3. There was a strong trend in the present experiment for regression suggestions to be more effective with subjects who had received a hypnotic induction. This trend could have been produced by any one or a combination of the many variables that were included in

TABLE 3

MEAN SCORES AND ANALYSIS OF VARIANCE OF POSTEXPERIMENTAL SUBJECTIVE REPORTS

	Experimental treatments			Average
	Hypnotic induction, eyes closed	No induction, eyes closed	No induction, eyes open	
Conditions of recall				
Recall without suggestions	1.0 (0.6) [0.5]	1.2 (1.0) [0.8]	1.8 (1.2) [0.8]	1.3 (0.9) [0.7]
Recall with motivational suggestions	1.3 (0.5) [0.5]	1.2 (1.0) [0.9]	1.8 (1.4) [1.3]	1.4 (1.0) [0.9]
Recall with regression suggestions	1.3 (1.1) [0.6]	1.3 (0.8) [0.4]	1.2 (1.4) [0.7]	1.3 (1.1) [0.6]
Average	1.2 (0.7) [0.5]	1.2 (0.9) [0.7]	1.6 (1.3) [0.9]	

Source of variance	df	MS	F
Experimental treatments (A)	2	1.34 (2.80) [1.21]	1.55 (3.04) [1.32]
Conditions of recall (B)	2	.21 (.23) [.84]	
A × B	4	.57 (.53) [.46]	
Error	81	.86 (.92) [.91]	
Total	89		

Note.—Scores on the *Imagined* item are listed first, scores on the *Felt* item are listed next in parenthesis, and scores on the *Believed* item are listed last in brackets.

the hypnotic induction procedure, for instance, by defining the situation to the subject explicitly as "hypnosis" and implicitly as a situation in which high response to suggestions was expected, or by suggesting to the subject that he would find it easy to respond to further suggestions (cf. Barber, 1965b; Barber & Calverley, 1964c). Further studies are needed to determine whether, with larger subject samples, suggestions to regress are significantly more effective when given with, rather than without a preceding hypnotic induction. If a significant effect is found, further research would then be necessary to ascertain which of the many variables included in a hypnotic induction procedure are instrumental and which extraneous to producing the effect.

An additional question needs to be asked with respect to the present experiment, namely, was regression related to recall? To answer this question, product-moment correlations were computed separately for Groups 3, 6, and 9 between their Recall scores and their Regression scores. (The Regression scores were computed by assigning 1 point for each of the three questions—for example, What is the month and the week?—that the subject answered in the regressed manner.) The correlations were .00, $-.68$ ($p < .05$), and .22, respectively, for Groups 3, 6, and 9. These coefficients indicate that, with respect to Groups 3 and 9, Regression scores were not related to Recall scores and that, with respect to Group 6, the more the subjects met criteria for regression, the *fewer* the number of nonsense syllables they correctly recalled. Stated in more general terms, these correlations indicate that subjects who meet criteria for regression, that is, who state that it is now a previous time, do not manifest enhanced recall for material they had learned at the previous time.

Postexperimental Subjective Reports

After it was stated that the experiment was over, all subjects were given a questionnaire to answer concerning the extent to which, when they were recalling the nonsense syllables during the experiment, they had (*a*) imagined, (*b*) felt, and (*c*) believed that they were back at the time when they had originally learned the syllables. It was expected that subjects given suggestions to regress would score markedly higher on each of these three questions than subjects not given such suggestions. The mean scores on the three postexperimental questions, together with analyses of variance of the scores, are presented in Table 3. This table shows that, contrary to expectation, as many subjects who were *not* given suggestions to regress as were, stated postexperimentally that, when they were recalling the syllables during the experiment, they had imagined, felt, and believed that they were back in Mr. Ryan's class, that it was the first week in October, and that they had just learned the nonsense syllables.

With respect to the *Imagined* item (Table 3, first score in each cell), the majority of subjects in each of the nine experimental groups stated that, during the experimental recall period, they had imagined that they were back in the original learning situation either "to some extent" (scores of 1) or "to a great extent" (scores of 2). Sixteen subjects (18%) testified that they "thoroughly imagined" that they were in the original learning situation (scores of 3); these 16 subjects were distributed more or less evenly over the nine experimental groups.

With respect to the *Felt* item (Table 3, scores in parentheses), subjects in most of the experimental groups stated on the average that they felt they were back in the original learning situation "to some extent" (scores of 1). With respect to the *Believed* item (Table 3, scores in brackets), in each of the experimental groups, about as many subjects testified that they had not believed that they were back at the previous time (scores of zero) as stated that they had believed that they were back "to some extent" (scores of 1). Eleven percent and 10% of the subjects, respectively, stated that they had "thoroughly felt" and "thoroughly believed" that they were back in the original learning situation; these subjects were scattered more or less equally over the nine experimental groups.

Unstructured interviews conducted with the subjects after they had answered the Post-Experimental Subjective Reports questionnaire provided a tentative explanation for these surprising results. The interview data indicated that, when attempting to recall the nonsense syllables during the experiment, most subjects in each of the nine experimental groups tried to place themselves in imagination in the original learning situation, that is, they tried to visualize the instructor, the list of nonsense syllables, and themselves learning and writing down the list. Since the subjects had typically attempted to enhance their experimental recollection of the syllables by trying to place themselves in fantasy in the original learning class, they were able to testify postexperimentally, more or less truthfully, that, during the experiment, they had to some degree imagined, felt, or believed that they were back in the original learning situation.

Three additional statements should be made with respect to the imagined, felt, and believed items which comprised the Post-Experimental Subjective Reports questionnaire:

1. These items were significantly intercorrelated, but the correlations were not as high as one might expect. The overall product-moment correlations were .47, .39, and .65, respectively, between imagined and felt, imagined and believed, and felt and believed.

2. Whether or not the subjects stated postexperimentally that, when recalling the syllables, they had imagined, felt, or believed that they were back in the original learning situation, was not related to the number of syllables that they had correctly recalled. The overall correlations between the Recall scores and the imagined, felt, and believed items were .07, −.05, and .13, respectively.

3. The degree to which the subjects in Groups 3, 6, and 9 met criteria for regression during the experiment was not related to their postexperimental testimony. The product-moment correlations between the Regression scores and the imagined, felt, and believed items of the Post-Experimental questionnaire were in no case significantly different from zero.

The discrepancies between the Regression scores and the scores on the Post-Experimental Subjective Reports questionnaire merit further comment. The reader will recall that all subjects in Group 3, 60% in Group 6, and 80% in Group 9 answered each of the three questions asked during the experimental regression period (for example, What is the month and the week?) as if they were regressed, that is, as if they were back in the original learning class. However, a far smaller percentage of subjects in each of these groups testified postexperimentally that, during the experimental recall period, they had "thoroughly" imagined, felt, or believed that they were back in the original learning situation. The discrepancies between the subjects' experimental and postexperimental statements may be due to various factors, such as (*a*) the questions asked during the experimental regression period were not worded in the same way as the questions asked postexperimentally, and (*b*) the experimental regression questions were answered orally whereas the postexperimental questions were answered by writing. Although these and other factors may have played a role, it seems that the most important factor in producing the differences in the experimental and postexperimental reports was as follows: The questions asked during the experimental regression period followed immediately upon repeated emphatic suggestions that the subject was in the original learning situation, for example, ". . . It is the first week in October, you are in Mr. Ryan's class . . . you have just learned the nonsense syllables. . . ." These suggestions carried the strong connotation that if the subject were to state that it was *not* the first week in October, that he was *not* in Mr. Ryan's class, etc., then he would be behaving in a negativistic manner characteristic of "poor" subjects, and he would be subverting the experiment and disappointing the experimenter. In contrast, the postexperimental questions were answered after it was stated that the experiment was over and without any explicit suggestions as to how the subject was to answer. Whatever suggestive factors were present in the postexperimental period were thus far less obvious, and presumably far less potent, than the explicit, direct, suggestive influence that was exerted during the experimental regression period.

The results obtained with the Post-Experimental Subjective Reports questionnaire raise a serious question pertaining to previous studies concerned with "hypnotic age regression." Prior studies usually included among the criteria for regression the hypnotic subject's postexperimental testimony that, during the experiment, he felt or believed that he had returned to an earlier chronological age. However, nonhypnotic subjects were *never* asked postexperimentally if, during the experiment, they too had felt or believed that they had returned to a previous age (see Barber, 1962b, for a review of this literature). The results of the present investigation strongly indicate that no conclusions concerning the effectiveness of hypnotic induction (or "hypnosis") in producing age regression can be deduced from any previous experiment in this area until the experiment has been replicated with the addition of a nonhypnotic control group. Subjects in the control group should be instructed to recall events from a previous time and should be asked postexperimentally whether, when recalling the events, they had felt or believed that they were back in the previous time. If subjects are assigned *at random* to hypnotic and nonhypnotic groups, investigators may be surprised to find that waking control subjects *who did not receive suggestions to regress* give postexperimental subjective reports which are very similar to those given by hypnotic subjects who received regression suggestions.

Differences in Performance of "Good" and "Poor" Hypnotic Subjects

At the close of the experiment, the subjects who had been exposed to a hypnotic induction (Groups 1, 2, and 3) were given additional suggestions of relaxation, drowsiness, and sleep and then were assessed on response to the Barber Suggestibility Scale (BSS). In accordance with previously established norms (Barber, 1965c), subjects obtaining Objective scores of 0–2, 2–4, 4–6, and 6–8 on the BSS were classified as poor, fair, good, and excellent hypnotic subjects, respectively. Two-thirds of the hypnotic subjects were rated as either excellent or good. Specifically, of the 30 subjects in the three hypnotic groups, 4 (13%) were rated as poor, 6 (20%) as fair, 11 (37%) as good, and 9 (30%) as excellent.[5]

Was response to the BSS related to proficiency of recall of the nonsense syllables

---

[5] Investigators who use the concept "depth of hypnosis" would most likely have categorized the excellent subjects as "deeply hypnotized," the good subjects as in "medium hypnosis," etc. When the construct "hypnotic depth" is used in this manner, however, it says no more than that subjects who had received a hypnotic induction procedure showed a certain degree of suggestibility and it fails to take into account the possibility that the subjects may have shown the same degree of suggestibility without a hypnotic induction (Barber, 1964; Barber & Calverley, 1965).

(Recall scores) or to the imagined, felt, or believed items that comprised the Post-Experimental Subjective Reports questionnaire? To answer this question, Pearsonian correlations were computed separately for Groups 1, 2, and 3 between scores on the BSS and Recall scores and between scores on the BSS and scores on the imagined, felt, and believed items. None of the correlations between the BSS and Recall scores were significantly different from zero. However, response to the BSS was consistently related to response to the imagined, felt, and believed items. Overall, scores on the BSS were correlated .36 with the imagined item, .54 with the felt item, and .57 with the believed item ($ps < .05$).

With respect to the hypnotic group that had received suggestions to regress (Group 3), there was no relationship ($r = .00$) between scores on the BSS and Regression scores.

In summary, these results indicate that: (*a*) The better hypnotic subjects (those obtaining higher scores on the BSS) did not recall more of the nonsense syllables than the poorer hypnotic subjects. (*b*) The better hypnotic subjects did not differ from the others in response to suggestions to regress. (Irrespective of their scores on the BSS, all hypnotic subjects in Group 3 who received regression suggestions responded to questions [e.g., "What is the month and the week?"] as if they were regressed.) (*c*) The better hypnotic subjects more often than the poorer subjects testified postexperimentally that, during the experimental recall period, they had imagined, felt, and believed that they were back in the original learning situation.[6]

REFERENCES

ÅS, A. The recovery of forgotten language knowledge through hypnotic age regression: A case report. *American Journal of Clinical Hypnosis,* 1962, **5**, 24–29.

BARBER, T. X. Experimental evidence for a theory of hypnotic behavior: II. Experimental controls in hypnotic age-regression. *International Journal of Clinical and Experimental Hypnosis,* 1961, **9**, 181–193.

BARBER, T. X. Experimental controls and the phenomena of "hypnosis": A critique of hypnotic research methodology. *Journal of Nervous and Mental Disease,* 1962, **134**, 493–505. (a)

BARBER, T. X. Hypnotic age-regression: A critical review. *Psychosomatic Medicine,* 1962, **24**, 286–299. (b)

BARBER, T. X. "Hypnosis" as a causal variable in present-day psychology: A critical analysis. *Psychological Reports,* 1964, **14**, 839–842.

BARBER, T. X. Effects of "hypnosis" on learning and recall: A methodological critique. *Journal of Clinical Psychology,* 1965, **21**, 19–25. (a)

BARBER, T. X. Experimental analyses of "hypnotic" behavior: A review of recent empirical findings. *Journal of Abnormal Psychology,* 1965, **70**, 132–154. (b)

BARBER, T. X. Measuring "hypnotic-like" suggestibility with and without "hypnotic induction"; psychometric properties, norms, and variables influencing response to the Barber Suggestibility Scale (BSS). *Psychological Reports,* 1965, **16** (Monogr. Suppl. 3—V16), 809–844. (c)

BARBER, T. X., & CALVERLEY, D. S. "Hypnotic-like" suggestibility in children and adults. *Journal of Abnormal and Social Psychology,* 1963, **66**, 589–597.

BARBER, T. X., & CALVERLEY, D. S. Comparative effects on "hypnotic-like" suggestibility of recorded and spoken suggestions. *Journal of Consulting Psychology,* 1964, **28**, 384. (a)

BARBER, T. X., & CALVERLEY, D. S. Effect of *E*'s tone of voice on "hypnotic-like" suggestibility. *Psychological Reports,* 1964, **15**, 139–144. (b)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of defining the situation as hypnosis and defining response to suggestions as easy. *Journal of Abnormal and Social Psychology,* 1964, **68**, 585–592. (c)

BARBER, T. X., & CALVERLEY, D. S. Empirical evidence for a theory of hypnotic behavior: Effects on suggestibility of five variables typically included in hypnotic induction procedures. *Journal of Consulting Psychology,* 1965, **29**, 98–107.

---

[6] If, in accord with many previous experiments, we had nonrandomly allocated good hypnotic subjects to the hypnotic treatments and poor hypnotic subjects or unselected subjects to the nonhypnotic treatments, would we have obtained a different set of results? The data presented above indicate that, if we had used this nonrandom method to allocate subjects, we still would have found that the hypnotic and nonhypnotic groups did not differ significantly in Recall scores and in Regression scores although they may have differed on the imagined, felt, and believed items (Post-Experimental Subjective Reports). However, as pointed out in the methodological introduction to this paper, if we had allocated good subjects to the hypnotic treatment but not to the nonhypnotic treatment, it would have been impossible to state whether any differences obtained between hypnotic and nonhypnotic groups on Post-Experimental Subjective Reports were due to the presence of "hypnosis" in one group and its absence in the other or to one group but not the other consisting of subjects who were highly suggestible with or without "hypnosis."

BARBER, T. X., KARACAN, I., & CALVERLEY, D. S. "Hypnotizability" and suggestibility in chronic schizophrenics. *Archives of General Psychiatry,* 1964, **11**, 439–451.

GLAZE, J. A. The association value of nonsense syllables. *Journal of Genetic Psychology,* 1928, **35**, 255–267.

HAMMER, E. F. Post-hypnotic suggestion and test performance. *Journal of Clinical and Experimental Hypnosis,* 1954, **2**, 178–185.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.

LEONARD, J. R. An investigation of hypnotic age-regression. Unpublished doctoral dissertation, University of Kentucky, 1963.

LYON-JAMES, S. C. The effect of posthypnotic suggestion upon the study and recall of written material. Unpublished doctoral dissertation, University of Denver, 1957.

MARCUSE, F. L. *Hypnosis: Fact and fiction.* Baltimore: Penguin, 1959.

PATTIE, F. A. A report on attempts to produce uniocular blindness by hypnotic suggestions. *British Journal of Medical Psychology,* 1935, **15**, 230–241.

ROSENTHAL, B. G. Hypnotic recall of material learned under anxiety and non-anxiety producing conditions. *Journal of Experimental Psychology,* 1944, **34**, 369–389.

SCHRENCK-NOTZING, A. F. VON. Ein experimenteller und kritischer Beitrag zur Frage des suggestiven Hervorrufung circumscripter vasomotorischer Veranderungen auf der ausseren Haut. *Zeitschrift Hypnotismus,* 1896, **4**, 209.

STALNAKER, J. M., & RIDDLE, E. E. The effect of hypnosis on long delayed recall. *Journal of General Psychology,* 1932, **6**, 429–440.

SUTCLIFFE, J. P. "Credulous" and "sceptical" views of hypnotic phenomena: A review of certain evidence and methodology. *International Journal of Clinical and Experimental Hypnosis,* 1960, **8**, 73–101.

TROFFER, S., & TART, C. Experimenter bias in hypnotist performance. *Science,* 1964, **145**, 1330–1331.

TRUE, R. M. Experimental control in hypnotic age regression states. *Science,* 1949, **110**, 583–584.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford hypnotic susceptibility scale.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

WHITE, R. W., FOX, G. F., & HARRIS, W. W. Hypnotic hypermnesia for recently learned material. *Journal of Abnormal and Social Psychology,* 1940, **35**, 88–103.

WOLBERG, L. R. *Medical hypnosis: I. The principles of hypnotherapy.* New York: Grune & Stratton, 1948.

(Received April 23, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 181–188

AN EMPIRICAL TEST OF THE FEMALE CASTRATION COMPLEX[1]

RACHEL B. LEVIN

Veterans Administration Hospital and the State University of New York Upstate Medical Center, Syracuse, New York

This study investigated the psychoanalytic hypothesis that intensity of the female castration complex (FCC) is greater in women with a masculine social role than in those with a feminine social role. Ss were normal, college-educated women, ages 30–55, divided into 2 groups: 26 career women (CW), unmarried, and in masculine occupations; and 25 homemakers, married, with 2 or more children, and not employed outside the home. The CW scored significantly higher on an overall measure of the FCC based on the Rorschach. The hypothesis was thus supported. The CW also scored significantly higher on 3 of the 8 individual components of the FCC (Penis Envy, High Activity, and High Need for Achievement). Some theoretical implications are discussed.

The female castration complex (also called, among other things, "penis envy" and the "masculinity complex") has been offered by psychoanalytic theorists as a major explanatory concept in the psychology of women. In general terms, the complex refers to the woman's presumed discontent with her femininity and her modes of reacting to this discontent. The complex is seen as occurring in varying degrees in most women, along a continuum from the normal to the abnormal; as being a conspicuous feature of feminine neuroses and pathological character formations; and as motivating a wide variety of behavioral manifestations. As an intervening variable, the female castration complex is supposedly related to observable clinical phenomena such as the feeling of having been castrated, envy of the penis, the desire to castrate men, or the wish-fulfilling fantasy of possessing a penis. In addition to, or in place of, these sexual manifestations, the complex may also be related to more general social attitudes and behavior, such as a derogatory attitude toward women and their traditional social role, envious hostility toward men, and the attempt to emulate the masculine social role. Two frequently described clinical phenomena that cut across the sexual-nonsexual differentiation are "wish-fulfillment" and "revenge" behavior; while both may occur in the same individual, one generally predominates. The wish-fulfillment type of behavior is illustrated by homosexuality, the pursuit of an intellectual or professional career, the avoidance of marriage and/or motherhood, hypomanic activity in any sphere, and gynecological disturbances of a functional nature. Some examples of the revenge type are frigidity, vaginismus, and prostitution (Abraham, 1920; Deutsch, 1944; Fenichel, 1945; Freud, 1933; Menninger, 1939; Thompson, 1943).

In spite of the broad agreement that psychoanalytic writers have shown as to these clinical or descriptive aspects of the complex, there have been many theoretical differences with regard to its genesis and basic nature. The complex is conceded by all theorists to be related to both sexual and nonsexual behavior, but there remains an unresolved issue as to which of these two is primary, and which is secondary or derived. According to the biological approach (Abraham, 1920; Deutsch, 1944; Freud, 1925, 1931, 1933; Hayward, 1943), the origin of the complex is literal envy of the male anatomy. It is presumed that during the phallic stage every little girl falls victim to this envy, which in the course of normal feminine development is eventually resolved by being transformed into the wish for a child (equivalent to penis) from the father (the Oedipus complex). If, however, the girl's bisexual disposition is weighted in the direction of masculinity, her castration complex remains unresolved. In the adult woman the complex may take "rationalized" forms such as protests against the greater social freedom of men, or the pursuit of an intellectual or professional career, but its genetic roots and underlying motives are literally sexual and biological.

---

[1] This article is based upon a dissertation submitted in partial fulfillment of the requirements for the PhD degree at Syracuse University. The author gratefully acknowledges the generous assistance of several members of the department, in particular Edward J. Murray, committee chairman, and Howard Friedman.

The culturally oriented theorists, on the other hand (Horney, 1937, 1939; Thompson, 1942, 1943), see the complex as stemming from social and family reactions to sexual differences. In the normal form in which it is believed to occur in most women in our culture, the complex is thought to represent a realistic wish for social equality with men. The complex may also be related to neurotic problems, such as deep feelings of inadequacy as a human being, which have become closely linked with the woman's sense of femininity. These problems arise from disturbed family relationships in which sexual differences play an important role, for example, the parents' preference for a male sibling. The complex may be symbolized in sexual forms such as literal penis envy, but its primary motives are nonsexual, for example, the needs for affection and self-esteem.

Notwithstanding the theoretical importance of the female castration complex, empirical studies of it have been few, both in the child and the adult (e.g., Blum, 1949; Eisenbud, 1951; Friedman, 1952; Landis, 1940; Levy, 1940; Schwartz, 1956). Most of the research has been characterized by methodological inadequacies such as the use of direct questions or the failure to demonstrate scorer reliability; and/or by conceptual oversimplification, such as the tendency to equate the female castration complex with the male castration complex (both often vaguely defined as "castration anxiety"). Other relevant studies have focused on limited areas such as masculine identification (e.g., Franck & Rosen, 1949), or have been more directly concerned with other topics (e.g., Bloomingdale, 1953).

In sum, there has been no systematic attempt to formulate a broad definition of the intervening variable of the female castration complex that would include its varied forms of expression in the adult woman; to design a reliable projective instrument for measuring these varied aspects or components; and, using this instrument, to test the validity of the construct by investigating its relationship to an independent, theoretically derived criterion. These were the general goals of the present research.

Since the complex has been assumed to be related to both literally sexual and more generalized nonsexual manifestations, it was decided to design a projective measure aimed at tapping both of these forms of expression. No hypothesis was formulated, however, with regard to the theoretical issue of the relative importance of the sexual as opposed to the nonsexual factors, since there was no way (evident to the author) of determining objectively which type of factor was basic and which was derived. In other words, sexual manifestations of the complex on the projective test might be symbolic expressions of generalized social attitudes (in the concrete language of the body), and conversely, nonsexual projective signs might be symbolic or derived representations of basically sexual factors. Our goal, then, was not to test the biological or cultural positions, but, very simply, to find out whether a group of women showing a type of overt behavior (the independent variable) theoretically related to the female castration complex, would show higher scores on a projective measure of the complex (the dependent variable) than a group of control women. The projective instrument was not designed to measure only the particular motivational factors that might be related to the specific behavioral criterion selected. It was devised with the more general aim of evoking as comprehensively as possible all the varied impulses and feelings, both sexual and nonsexual, both wish-fulfillment and revenge, that are common to most of the accounts of the complex.

The specific behavioral criterion that was selected was masculinity of social role or style of life. In broad terms, analytic theory proposes that the intensity of the woman's castration complex, an intervening motivational variable, is related to the degree of masculinity of her social role, an overt behavioral variable. The research hypothesis, derived from this general proposition, was as follows: *Intensity of the female castration complex as measured by a projective test is significantly* *greater in women with a masculine social role than in women with a feminine social role.* Of course, masculinity of social role is only one of the many forms of behavior to which the complex has been theoretically related; this particular behavioral variable was selected on the basis of expediency, namely—the relative ease of obtaining research subjects. The hypothesis was not intended to imply that an overtly masculine style of life was the only way in which the castration complex expressed itself.

METHOD

Subjects. The research subjects consisted of 51 white, native-born, college-educated American women between the ages of 30 and 55, in good physical health, and with no history of psychiatric treatment. These subjects were selected from a pool of 499 women, local alumnae of various colleges, who responded to an innocuous questionnaire entitled "A Socio-Psychological Survey of American Women." The 51 women were divided into two groups: (*a*) 26 career women, who represented a masculine social role, and (*b*) 25 homemakers, who represented a feminine social role. Two criteria were used in determining masculinity or femininity of social role: (*a*) family status (marriage and children), and (*b*) occupation. The two research groups contrasted clearly with respect to both of these criteria. Thus, the career women were unmarried and employed in masculine occupations, while the homemakers were married, living with their husbands, mothers of two or more children, and not gainfully employed outside the home. Masculine occupations were defined by a high ratio of men to women in census statistics, or by a high level of status and responsibility; using these definitions, two psychologists showed strong agreement in assigning high ranks on masculinity of occupation to the career women ($\rho = .92$, $p < .001$). They included lawyers, physicians, research scientists, business executives, and university faculty members; 18 of them had higher degrees, including 9 doctorates. The homemakers had a median of 4 children each; none of them had higher degrees, although they were all college graduates.

Later, the two groups were compared with respect to Vocabulary subtest scores on the Wechsler Adult Intelligence Scale (WAIS); socioeconomic status of family of origin, as measured mainly by father's education and occupation; and current socioeconomic status, as measured by the education and occupation of the subject herself in the case of the career women, and of the subject's husband in the case of the homemakers. Comparison by means of the chi-square test showed no significant differences between the two groups on any of these variables. In both groups, the median age was in the range of 35–39, and the majority were Protestant. The median Verbal IQ, as estimated by the WAIS Vocabulary scores, was 129 for the career women and 132 for the homemakers.

Procedure. All subjects were tested by the investigator in individual appointments that lasted about 2 hours each. Standard clinical procedures were used in administering first, the Rorschach test, and then, the Vocabulary subtest of the WAIS. These were followed by two especially designed instruments that the subjects filled out by themselves: an opinion questionnaire focusing on the social role of women (not reported on here), and a biographical questionnaire (used in the final equating of the groups). Cooperation by the subjects was excellent.

The female castration complex measure (FCCM). The first step in constructing the measuring instrument was to delineate the major components (or modes of expression) of the female castration complex. On the basis of a survey of the psychoanalytic literature, 8 such specific components were selected (see Table 1). While these 8 were not assumed to be a comprehensive list of "pure" factors, they were felt to include most of the main features of the complex as described in the literature. The Rorschach was chosen as the most appropriate test upon which to base the measuring instrument. For each of the 8 components, a corresponding Rorschach measure consisting of one or more signs was formulated (Table 1); the 8 Rorschach measures together formed the FCCM. The individual signs were derived from the Rorschach literature and from common clinical usage with this test. They included both traditional Rorschach scores (such as form level) and thematic and stylistic features of the protocols that conventionally do not receive formal scoring. While these signs might, of course, reflect factors other than the female castration complex, they were assumed to have some degree of face validity as indicators of the various aspects of the complex as they are described in Table 1. For a detailed description of the FCCM, including the rationale for each of the signs, see Levin (1962).[2]

Since some of the FCCM signs consisted of conventional Rorschach scores, one person first scored the Rorschach protocols in the usual way. Then another person scored the protocols on the FCCM by deciding which, if any, FCCM signs were present in each Rorschach response. The FCCM scorer had first been trained in the application of the FCCM by the use of five protocols not included in the study. Neither of the two scorers had any knowledge of the design of the research. On 10 protocols (5 from each group of subjects) the first scorer (conventional Rorschach scores) showed percentages of agreement

---

[2] The complete manual of the Rorschach Female Castration Complex Measure, the instructions for its use, and the description of its rationale, have been deposited with the American Documentation Institute. Order Document No. 8532 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $2.75 for microfilm or $7.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1

THE EIGHT COMPONENTS AND CORRESPONDING RORSCHACH MEASURES OF THE FEMALE CASTRATION COMPLEX

Component	Rorschach measure[a]
1. *Castration Anxiety:* The woman's fear and rejection of her femininity in a specifically genital sense. Conception of the female genitals as castrated, damaged, and inferior.	1. *Female Sexual Disturbance:* Disturbance of structure, theme, or style in female sex (genital) percepts or in responses to popular female sex areas. For example, "Blood stain" in II D3.
2. *Penis Envy:* The woman's envy of the male genitals, in a literal sense. Wish to castrate men in revenge for her own castration. Projection of her hostility with consequent fear of the penis.	2. *Male Sexual Distrubance:* Disturbance of structure, theme, or style in male sex (genital) percepts or in responses to popular male sex areas. For example, any percept scored F- in II D4.
3. *Confused Body Image:* The woman's confusion about the male and female body images, stemming from her own ambivalent sexual identification. Blurring or denial of the physical distinctions between the sexes.	3. *Difficulty in Assigning Sex:* Avoidance, confusion, or misperception in assigning sex to human or humanlike percepts. For example, "Two people, I can't tell if they're men or women."
4. *Penis Fantasy:* The woman's defense against her castration by denial and the wish-fulfilling fantasy of possessing a penis, in a literal sense.	4. *Phallic Symbolism:* Explicit reference to phallic characteristics of percepts. For example, "Person with a very long nose."
5. *Rejection of Femininity:* The woman's fear and rejection of her femininity in a general (not specifically sexual) sense. Image of women as weak, inferior, and victimized, or defensive conception of them as powerful and dangerous.	5. *Female General Disturbance:* Disturbance of structure, theme, or style in female general (nonsexual) percepts or in responses to popular female general areas. For example, any percept scored F- in I D4.
6. *Envy of Men:* The woman's envy of men in a general (not specifically sexual) sense. Conception of men as superior and powerful, or defensive image of them as inferior and weak. Projection of her hostility with consequent fear of men.	6. *Male General Disturbance:* Disturbance of structure, theme, or style in male general (nonsexual) percepts or in responses to popular male general areas. For example, "Scowling man."
7. *High Activity:* The woman's identification with "masculine activity." An active, energetic orientation to the world, and strong needs for independence and self-assertion.	7. *Active Human Movement:* Perception of human or humanlike percepts in active kinds of movement. For example, "Clowns doing tricks."
8. *High Need for Achievement:* The woman's identification with "masculine striving." Competitiveness, ambition, and strong needs for status and achievement, especially in masculine fields.	8. *Achievement Content:* Responses with content reflecting strong drives for achievement and prestige, especially in masculine fields. For example, "An emblem."

[a] For the derivation of the popular areas of Measures 1, 2, 5, and 6, see Levin (1962). Locations are those listed by Beck (1944).

with another psychologist that ranged from 84 to 91. The FCCM scorer's percentage of agreement with a second psychologist on 20 protocols (10 from each group) was 91.

Nine main FCCM scores were computed for each subject: an overall FCCM score and eight specific measure scores. General Rorschach productivity (length of responses and total number of responses), was controlled in these main scores by counting each response a maximum of once for each FCCM measure (even if it contained more than one sign of that measure), and by expressing each FCCM score as a percentage of the subject's total number of responses (Total R). Thus, the subject's main overall FCCM score consisted of the number of her responses containing one or more FCCM signs, divided by her Total R, and multiplied by 100; and each of her eight main measure scores consisted of the number of her responses containing one or more signs of that measure, divided by Total R, and multiplied by 100. In addition, a second set of scores was computed for six measures, controlling productivity in the special categories of Rorschach responses that were pertinent to those measures, that is, the categories that were potential carriers of signs. For example, the pertinent category of responses for Measure 7, Active Human Movement (human or humanlike percepts in active kinds of movement), was the subject's total number of human movement (M) responses (both active and passive), since Measure 7 could occur only in M responses. Thus, the Measure-7 score controlling for productivity in the pertinent category of Rorschach responses consisted of the number of responses containing active M, divided by the total number of responses scored M, and multiplied by 100; whereas in the main Measure-7 score, the number of active M was divided by Total R. Scores controlling for FCCM-pertinent productivity were not necessary in the case of Measures 4 and 8, since these measures could occur in any responses. For each FCCM measure, all scores (51 subjects) were dichotomized at the median, and the chi-square value (corrected for continuity) of the difference between the groups was determined. For the sake of consistency, all tests of significance were one-tailed. In order to obtain a measure of the degree of relationship between the homemakers-career women dichotomy and the FCCM scores, Φ coefficients were computed for all chi-square values.

RESULTS

Main findings. The main findings were the results for the overall FCCM scores, the measure for the overall intensity of the female castration complex, taking all the specific scores into account. As Table 2 (scores controlling for Total R) indicates, the career women were significantly higher in overall FCCM scores ($p < .001$). A second analysis using the FCCM scores controlling productivity in FCCM-pertinent responses yielded similar results (see Table 3). The research hypothesis was thus supported.

Although both groups of subjects were psychologically "normal" as defined by the criterion of never having received psychiatric treatment, this was obviously a very crude criterion. It was possible, therefore, that the two groups differed widely in general disturbance or maladjustment, and that it was this difference that accounted for the overall FCCM difference (since many FCCM signs were by definition signs of disturbance). In order to investigate this possibility, a measure of general disturbance, or disturbance independent of the female castration complex, was derived by considering only those responses that did not contain any FCCM signs and/or were not in the categories of responses pertinent to the FCCM. The percentage of these non-FCCM-pertinent responses that contained the same structural signs of disturbance used in the FCCM (such as F-) was the score for general disturbance. Comparison of the two groups on these scores showed no significant difference ($\chi^2 = 1.62$, p *ns*). Also, the correlation between the overall FCCM scores and the general disturbance scores, although positive, was not significant ($\chi^2 = 2.38$, $\Phi = .22$, p *ns*). These findings do not exclude the possibility that disturbance of a general sort was manifesting itself only in the FCCM, especially since the measure of general disturbance, which included only structural signs, was probably not as sensitive as the FCCM, which contained also stylistic and thematic signs. Nevertheless, the findings with regard to general disturbance do suggest that the significant differences between the two groups on the FCCM probably could not be attributed to differences in overall level of adjustment.

TABLE 2

RELATIONSHIP BETWEEN THE HOMEMAKERS-CAREER WOMEN DICHOTOMY AND THE INDIVIDUAL FCCM MEASURES: SCORES CONTROLLING GENERAL RORSCHACH PRODUCTIVITY

Measure	Median		χ^2	Φ	p[a]
	Homemakers	Career women			
1. Female sex disturbance	0	0	.04	.03	*ns*
2. Male sex disturbance	1.8	5.4	3.36	.26	<.05
3. Difficulty assigning sex	6.3	5.3	.02	.02	*ns*
4. Phallic symbolism	4.0	6.6	2.38	.22	*ns*
5. Female general disturbance	5.3	8.9	4.43	.30	<.05
6. Male general disturbance	9.1	13.0	.97	.14	*ns*
7. Active human movement	7.1	13.2	4.43	.30	<.05
8. Achievement content	3.2	8.9	7.10	.37	<.01
Overall FCCM	28.0	50.0	18.88	.61	<.001

[a] One-tailed tests.

TABLE 3

RELATIONSHIP BETWEEN THE HOMEMAKERS-CAREER WOMEN DICHOTOMY AND THE INDIVIDUAL FCCM MEASURES: SCORES CONTROLLING FCCM-PERTINENT RORSCHACH PRODUCTIVITY

Measure	Median		χ^2	Φ	p[a]
	Homemakers	Career women			
1. Femal sex disturbance	0	0	.04	.03	*ns*
2. Male sex disturbance	25	100	5.66	.33	<.01
3. Difficulty assigning sex	50	32	.95	−.14	*ns*
5. Female general disturbance	67	93	2.38	.22	*ns*
6. Male general disturbance	75	100	2.38	.22	*ns*
7. Active human movement	50	75	8.93	.42	<.01
Overall FCCM	80	122	19.15	.61	<.001

Note.—Measures 4 and 8 were not affected by FCCM-pertinent productivity.

[a] One-tailed tests.

Analysis of specific measures. When the eight individual measures were intercorrelated, only one of the resulting 28 intercorrelations (that between Difficulty in Assigning Sex and Male Sex Disturbance) was significant in the predicted (positive) direction ($\Phi = .25$, $p < .05$). Since one significant correlation out of a total of 28 could be due to chance, it was concluded that the eight measures were independent of each other.

Of the eight individual measures, four were significantly higher in the career women when the main scores (controlling for Total R) were used (see Table 2). The results were essentially the same when scores controlling for productivity in FCCM-pertinent responses were analyzed, except that one significant measure no longer differentiated significantly between the groups, while the chi-square values of two other significant measures increased (see Table 3). If only the more conservative findings are selected from both analyses, these results may be summarized as follows: three measures differentiated significantly between the groups, one at the .01 level (Achievement Content) and two at the .05 level (Male Sex Disturbance and Active Human Movement); five measures failed to differentiate significantly between the groups (Female Sex Disturbance, Difficulty in Assigning Sex, Phallic Symbolism, Female General Disturbance, and Male General Disturbance).

Analysis of structural, thematic, and stylistic signs. Since thematic and stylistic types of signs might be assumed to be more amenable to conscious control than structural ones, it was important to determine whether all three types of signs contributed to the significant findings. Therefore, within each of Measures 1, 2, 5, and 6, the signs within each type (structural, thematic, and stylistic, respectively) were combined, and each combination (with Total R controlled) was correlated with the homemakers-career women dichotomy. The mean correlation of the structural combinations over all four measures was then computed, and similarly, the mean correlations of the thematic and stylistic combinations. For each of these three types of combinations, the mean correlation with the homemakers-career women dichotomy was significant ($p < .05$). All three types of signs, then, contributed significantly to the results.

DISCUSSION

The major finding of this study was that a group of career women (not married, and in masculine occupations) showed significantly higher scores than a comparable group of homemakers (married, with children, and not employed) on a Rorschach measure for intensity of the female castration complex. On the assumption that the two groups of subjects represented two contrasting social roles or general styles of life, the one more masculine and the other more feminine, the study provides some support for the hypothesis that intensity of the female castration complex is significantly greater in women with a masculine social role than in those with a feminine social role. Also, some evidence is furnished for the validity of the Rorschach-based instrument that was developed to measure the complex.

In addition to the overall Rorschach measure of the complex, three of the eight individual components were also significantly higher in the group of career women: Penis Envy (envious, hostile, and/or fearful reactions to the male genitals); High Activity (energy, independence, and self-assertion); and High Need for Achievement (ambition and drive for status). In other words, the career women relative to the homemakers were found to be more active and assertive, more oriented towards status and achievement, and more disturbed in their reactions to male sexuality. These motivational variables seem quite consistent, in a common sense way, with the careers these women were engaged in as well as with their failure to marry.

The findings do not imply, of course, that a masculine social role is the only possible mode of expression of the female castration complex, or conversely, that all career women necessarily have a high intensity of the complex. The latter becomes clearly more doubtful, for example, in the case of career women who are married and/or working in feminine types of occupations. The ambition and vitality inherent in the pursuit of a career might, of course, be found in conflict-free femininity. Also, the failure to marry might stem from factors other than the castration complex, for example, overdependence upon a parent, neurotic inhibition, or a philosophical objection to the institution of marriage. Nevertheless, the study suggests that there is a relationship between some motivational variables that have been included in descriptions of the female castration complex, and a social role that involves remaining single and pursuing a masculine type of career.

The question might be asked, do the significant Rorschach differences between the two groups reflect factors that contributed to the different role choices, or consequences of those different social roles? The available data do not provide an answer to this question. It is possible, to mention only one example, that the career women failed to marry because of reasons irrelevant to the castration complex (e.g., rejection by the men of their choice), and then reacted with negative feelings about sex and a compensatory channeling of their energies into their careers (particularly since they were faced with the necessity of supporting themselves). This study points only to a relationship between certain motivational variables as measured by the Rorschach and the overt behavior of the research subjects.

Why did five of the eight components of the complex fail to differentiate between the career women and the homemakers? First, it is possible that the Rorschach measures of these five components were not as sensitive as the others, even though all eight Rorschach measures were derived in the same fashion. It is also possible that all the Rorschach measures were equally valid, but that no real differences obtained between the two groups of women with respect to the five nonsignificant components. That is, one might conclude that the masculine role behavior of the career women of the study (both with respect to their choice of occupation and their failure to become wives or mothers) is related only to the three significant components. Other groups of women such as homosexuals, women with functional gynecological disorders, discontented housewives, career women who are married or in feminine types of occupations, and neurotic or psychotic patients, might show higher scores on other specific measures of the complex.

This raises the question of whether the female castration complex can be viewed as a unitary variable. The absence of significant intercorrelations among the specific measures would argue against the idea of a unitary variable. If different aspects of the complex do not "hang together," but are correlated with different types of behavior, then the phenomena under discussion may be unrelated to each other. The results of this study, for example, could be explained simply in terms of differences between the two groups of women in heterosexual adjustment, activity level, and achievement drive, without invoking the castration complex as an explanatory concept. It is possible, on the other hand, that the absence of significant intercorrelations among the specific Rorschach measures might be due to the lack of precision of some or all of the measures. Further research would be necessary in order to investigate more fully whether and to what extent the various factors of the complex covary, and in order to determine how useful or parsimonious the construct of the female castration complex might be.

REFERENCES

Abraham, K. Manifestations of the female castration complex (1920). In *Selected papers of Karl Abraham*. New York: Basic Books, 1954. Pp. 338–369.

Beck, S. J. *Rorschach's Test*. Vol. I. New York: Grune & Stratton, 1944.

Bloomingdale, E. C. Psychological aspects of essential dysmenorrhea. Unpublished doctoral dissertation, Radcliffe Graduate School, 1953.

Blum, G. S. A study of the psychoanalytic theory of psychosexual development. *Genetic Psychology Monographs,* 1949, **39**, 3–99.

Deutsch, H. *The psychology of women*. Vol. I. New York: Grune & Stratton, 1944.

Eisenbud, R.-J. Factors influencing the repudiation of femininity: A comparison of professional and homemaking women. Unpublished doctoral dissertation, Radcliffe Graduate School, 1951.

Fenichel, O. *The psychoanalytic theory of neurosis.* New York: W. W. Norton, 1945.

Franck, K., & Rosen, E. A projective test of masculinity-femininity. *Journal of Consulting Psychology,* 1949, **13**, 247–256.

Freud, S. Some psychological consequences of the anatomical distinction between the sexes (1925).

In *Collected papers*. Vol. V. New York: Basic Books, 1959. Pp. 186–197.

FREUD, S. Female sexuality (1931). In *Collected papers*. Vol. V. New York: Basic Books, 1959. Pp. 252–272.

FREUD, S. The psychology of women. In *New introductory lectures on psychoanalysis*. New York: W. W. Norton, 1933. Pp. 153–185.

FRIEDMAN, S. An empirical study of the castration and Oedipus complexes. *Genetic Psychology Monographs*, 1952, **46**, 61–130.

HAYWARD, E. P. Types of female castration reaction. *Psychoanalytic Quarterly*, 1943, **12**, 45–66.

HORNEY, K. *The neurotic personality of our times*. New York: Norton, 1937.

HORNEY, K. Feminine psychology. In *New ways in psychoanalysis*. New York: W. W. Norton, 1939. Pp. 101–119.

LANDIS, C. *Sex in development*. New York: P. B. Hoeber, 1940.

LEVIN, R. B. The psychology of women: An empirical test of a psychoanalytic construct. Unpublished doctoral dissertation, Syracuse University, 1962.

LEVY, D. M. "Control-situation" studies of children's responses to the differences in genitalia. *American Journal of Orthopsychiatry*, 1940, **10**, 755–762.

MENNINGER, K. A. Somatic correlations with the unconscious repudiation of femininity in women. *Journal of Nervous and Mental Disease*, 1939, 89, 514–527.

SCHWARTZ, B. J. An empirical test of two Freudian hypotheses concerning castration anxiety. *Journal of Personality*, 1956, **24**, 318–327.

THOMPSON, C. Cultural pressures in the psychology of women. *Psychiatry*, 1942, **5**, 331–339.

THOMPSON, C. "Penis envy" in women. *Psychiatry*, 1943, **6**, 123–135.

(Received April 23, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 189–195

APPRAISAL OF SOCIAL MATURITY BY MEANS OF THE CPI

HARRISON G. GOUGH

University of California, Berkeley

The CPI is scaled for "folk concepts," dimensions of interpersonal behavior theorized to be culturally universal. Social maturity is a significant pattern of behavior which one should be able to assess from these variables. The optimum constellation of scales for this purpose includes socialization, responsibility, flexibility, and dominance, with positive weightings, and good impression and communality with negative. An equation composed of these variables validly differentiated delinquent from nondelinquent males in the United States and in Italy. Among nondelinquents, it distinguished between cheaters and noncheaters on course examinations, and between more- and less-responsible high school students. Adjectival analyses revealed high-scorers on the equation to be dependable, foresighted, and capable.

In selecting variables for scaling in the California Psychological Inventory (CPI; Gough, 1957), two principles were of central importance. The first was that the dimensions should possess the widest possible relevance for social and interpersonal behavior. Relevance here signifies extent or range of implications for situations in the life of any one individual, and also extent or range of applicability to different societies and cultures. Contextual validity, of the sort envisaged, would appear almost by definition to be embodied in the folk concept, that is, in those analytic and classificatory concepts which arise directly and immediately out of interpersonal living and which therefore may be presumed to be culturally universal.

The rationale for this choice of variables should be clear. If an inventory is intended for use in the study of interpersonal and interactional behavior, then a plausible focus of measurement would be on those concepts which are universally employed to describe and analyze such behavior. A first aim of the inventory, therefore, is to attempt to provide measures of a selected number of these folk dimensions of social behavior.

Its second emphasis is on those qualities of interpersonal disposition which are intrinsically favorable, or which may be hypothesized as important determiners of positive and socially constructive behavioral outcomes. In certain specific, and perhaps simple, instances the forecasting of these outcomes may be accomplished by a single scale or measure; more often, interpretation and prediction must rest on a pattern or combination of measures. Scholastic achievement, for example, may be predicted from the two need:achievement scales of the CPI, but a more accurate forecast may be derived from a constellation of six scales (Gough, 1964). In general, the more complex and multi-faceted the behavioral pattern which is to be forecast, the more likely will it be that a cluster or combination of variables will exceed the best that can be done with a single scale.

In this paper attention will be centered on one such pattern, that of social maturity. In every culture there is a process of social development, in which the individual may move from an untutored, egocentric status to a condition of interpersonal affiliation and harmony in the circumstances of social governance. No one begins fully estranged or alienated, and no one ends by being fully socialized or assimilated. But everyone, at each stage of his life, occupies an identifiable locus on this implicit scale of development, and the concept of social maturity is therefore universally meaningful.

The Socialization scale (Gough, 1960; Gough & Peterson, 1952) is directly relevant to this concept of social maturity, and one would expect to find it in any cluster on the inventory developed to assess the concept. Socialization, however, is not the same as social maturity. Socialization involves the internalization of rules, constraints, and values, with an emphasis on their preservation and maintenance.

Social maturity, in its highest form, involves the creation of new order, and hence the destruction of old order. The highly socialized individual can live by the rules, however oppressive, and in fact welcomes their challenge. The socially mature individual, on the other hand, although able to adapt to convention, is receptive to change and innovation, and under repressive conditions may set himself against the established order. The significance of this for measurement is that one should expect a combination of scales to do better than the So scale alone in identifying persons characterized by greater or lesser degrees of social maturity.

METHOD AND PROCEDURE

From the research files on the CPI, a large and heterogeneous sample of nondelinquent males ($N = 4{,}628$) was selected to represent average or above-average levels of social maturity. This sample was composed of 1,220 high school students, 2,000 college students, and 1,408 employed adult males including occupations such as electronic technician, machine operator, correctional officer, bank manager, city school official, psychiatric aide, research scientist, architect, and business executive.

TABLE 1

ORIGINAL SAMPLES: COMPARISON OF DELINQUENTS AND NONDELINQUENTS ON THE SCALES OF THE CPI

Scale	Nondelinquent[a]		Delinquent[b]		diff.	$r_{p.bis}$
	M	*SD*	*M*	*SD*		
Do	27.64	6.73	20.92	5 97	6.72*	+.42
Cs	19.55	4.61	14.91	4.42	4.64*	+.42
Sy	24.49	5.57	21.12	5.32	3.37*	+.27
Sp	35.82	6.31	31.98	5.84	3.84*	+.27
Sa	21.10	4.39	19.05	4.04	2.05*	+.21
Wb	36.61	5.12	31.87	6.31	4.74*	+.36
Re	30.34	5.50	20.76	5.90	9.58*	+.61
So	36.83	5.45	26.38	5.36	10.45*	+.66
Sc	28.58	7.92	24.91	7.75	3.67*	+.21
To	22.24	5.46	15.12	5.39	7.12*	+.51
Gi	17.64	6.55	16.47	6.06	1.17*	+.08
Cm	25.61	2.36	24.30	3.32	1.31*	+.22
Ac	26.96	5.49	20.48	5.42	6.48*	+.47
Ai	19.13	4.86	13.64	4.40	5.49*	+.47
Ie	38.66	6.00	30.83	6.65	7.83*	+.50
Py	11.12	3.02	8.86	2.89	2.26*	+.33
Fx	9.94	3.96	8.35	3.48	1.59*	+.19
Fe	16.12	3.63	16.26	3.61	−0.14	−.02

[a] $N = 2{,}146$.
[b] $N = 881$.
* $p < .01$.

To represent a lower level of social maturity, 1,290 institutionalized delinquent and adult prison inmates were chosen. About half of these came from Youth Authority agencies in California, and the remainder from adult centers in California and other states.

For purposes of analysis, an initial sample of 2,146 nondelinquents (proportionately drawn from the three categories) versus 881 delinquents was defined, and a cross-validating sample of 2,482 nondelinquents versus 409 delinquents. The delinquents and nondelinquents were compared on each of the 18 scales of the inventory and on two regression equations developed in the first sample and cross-validated in the second.

For cross-cultural validation, a sample of 659 nondelinquent Italian men was contrasted with a sample of 38 males in an Italian penal institution. To extend the range of implications of the equations, a sample of 78 college students who had been identified as cheaters or noncheaters on examinations (Hetherington & Feldman, 1964) was studied. Another validational sample consisted of 659 high-school males from five different cities. In each school the principal was asked to nominate three or four boys in each class who were very high on the trait of "responsibility," and an equal number who were very low. These nominations may be conceived of as a partial criterion of social maturity. Fifty boys were nominated as "highs" and 51 as "lows."

Finally, to specify the conceptual and psychological implications of the equations (cf. Gough, 1965), adjectival analyses were conducted in three samples: 100 military officers studied in personality assessment (MacKinnon, Crutchfield, Barron, Block, Gough, & Harris, 1958), and a fraternity (41 males) and a sorority (45 females) at the University of California.

FINDINGS

Table 1 presents statistical data derived from a comparison of delinquents and nondelinquents in the original sample. Seventeen of the 18 scales of the inventory differentiate significantly between the two groups, and in each instance the nondelinquents attain higher scores. If point-biserial correlations are used to indicate the magnitude of the relationships, the highest coefficient, as would be expected, is given by the Socialization scale with a coefficient of +.66; the next highest is Responsibility, with a coefficient of +.61, and the third highest is Tolerance, with a coefficient of +.51.

Two regression equations were next developed, using a stepwise multiple-correlation technique programed for the 7090 Computer

TABLE 2

CROSS-VALIDATING SAMPLES: COMPARISON OF DELINQUENT AND NONDELINQUENT MALES ON THE SCALES OF THE CPI, AND ON TWO REGRESSION EQUATIONS

Variable	Nondelinquents[a]		Delinquents[b]		diff.	$r_{p.\text{bis}}$
	M	*SD*	*M*	*SD*		
Do	27.93	6.62	21.18	5.78	6.75*	+.34
Cs	19.67	4.62	14.79	4.22	4.88*	+.35
Sy	24.88	5.31	21.35	5.37	3.53*	+.23
Sp	36.31	5.91	31.95	5.94	4.36*	+.25
Sa	21.32	4.07	19.16	4.61	2.16*	+.18
Wb	36.78	4.97	31.38	6.11	5.40*	+.34
Re	30.28	5.21	21.27	6.36	9.01*	+.50
So	36.88	5.32	27.51	5.37	9.37*	+.52
Sc	28.41	7.86	24.10	8.45	4.31*	+.19
To	22.28	5.48	14.68	5.50	7.60*	+.44
Gi	17.77	6.48	15.85	6.56	1.92*	+.10
Cm	25.69	2.19	24.23	3.22	1.46*	+.21
Ac	26.96	5.30	20.57	5.87	6.39*	+.38
Ai	19.12	4.86	13.53	4.89	5.59*	+.37
Ie	38.67	5.97	30.68	6.69	7.99*	+.42
Py	11.19	3.05	8.74	2.64	2.45*	+.27
Fx	9.91	3.85	8.26	3.56	1.65*	+.15
Fe	16.16	3.69	15.80	4.07	0.36	+.03
Equation 1	50.43	3.51	42.85	3.70	7.58*	+.60
Equation 2	50.76	3.52	42.59	3.53	8.17*	+.63

[a] $N = 2{,}482$.
[b] $N = 409$.
* $p < .01$.

at the University of California Computer Center.[1] In this analysis, the nondelinquents were assigned scores of 2 and the delinquents scores of 1. The first equation includes the optimum set of three variables yielded by the analysis; the number of variables was set at three so that this first equation would not be too cumbersome for hand calculation. The second equation includes six variables, and affords a more precise and accurate differentiation when a computing facility is available.

$$\text{Social Maturity } \#1 = 25.701 + .408\text{Re} + .478\text{So} - .296\text{Gi} \quad [1]$$

$$\text{Social Maturity } \#2 = 28.062 + .148\text{Do} + .334\text{Re} + .512\text{So} - .317\text{Gi} - .274\text{Cm} + .227\text{Fx} \quad [2]$$

The weights in the two equations are the computing weights, for use with raw scores on the scales, and the constants are set so that the average scores for each equation on an unselected sample of males will be 50. Interpretation of the equations will be deferred until cross-validating and other data have been presented, but the variables in each equation may be named at this point. Equation 1 includes the Responsibility and Socialization scales with positive weights, and the Good Impression scale with a negative weight; Equation 2 includes Dominance, Responsibility, Socialization, and Flexibility with positive weights, and Good Impression and Communality with negative loadings.

Cross-validation of the scales of the inventory and the two equations is offered in Table 2. As before, 17 of the 18 scales of the inventory differentiate significantly between the delinquent and nondelinquent groups, and in each instance the nondelinquents attain higher scores. The two equations also discriminate significantly between the two groups.

The point-biserial correlations in Table 2 are consistently lower than those in Table 1, the principal reason being the shift in proportion of cases in the delinquent and nondelinquent categories (29% delinquent, 71% nondelinquent in the original sample, and 14% versus 86% in the cross-validating sample); for a given mean difference and total variance, the magnitude of the point-biserial will decrease as the split of the sample departs from 50–50.

However, the ordering of the scales of the inventory is similar in both instances. In Table 2, the highest coefficient is found for the Socialization scale (+.52), the next highest for Responsibility (+.50), and the third highest for Tolerance (+.44). The two equations, as anticipated, exceed all of these values: for Equation 1 the point biserial was +.60, and for Equation 2 the coefficient was +.63.

CROSS-CULTURAL VALIDATION

The next step in the analysis was the cross-cultural validation. Scores on both equations were computed for a sample of 659 Italian males tested on the Italian edition of the Inventory,[2] and likewise for a sample of 38

---

[1] The writer wishes to thank the Computer Center for granting time for this study and Q. Welch, who conducted the analyses.

[2] Published in 1959 by the Organizzazioni Speciali, via R. Franchi, 5, Florence, Italy.

TABLE 3

DISTRIBUTIONS OF SCORES OF CROSS-VALIDATING AND CROSS-CULTURAL SAMPLES ON THE TWO REGRESSION EQUATIONS

Score	Equation 1				Equation 2			
	American		Italian		American		Italian	
	Non-del.	Del.	Non-del.	Del.	Non-del.	Del.	Non-del.	Del.
62					1			
61	1				1			
60	2				3			
59	7				10		1	
58	16				24		0	
57	41		1		53		0	
56	71		2		98		1	
55	107		5		134		2	
54	194		19		226		5	
53	270	2	30		265	2	11	
52	312	2	45		289	5	23	
51	301	9	60		303	4	67	
50	284	6	72		239	4	58	1
49	228	6	83		210	6	72	0
48	182	20	75	1	195	13	89	0
47	161	22	68	2	147	18	84	1
46	98	25	66	3	101	26	72	3
45	62	33	51	7	72	33	68	2
44	44	53	28	6	37	45	53	9
43	28	36	17	4	31	50	24	5
42	32	42	18	4	15	49	15	5
41	15	53	8	5	12	37	6	3
40	17	23	4	1	9	34	5	3
39	5	29	2	1	1	35	3	2
38	2	19	2	3	6	20		1
37	1	14	1	0		14		1
36	0	7		0		9		2
35	1	4		1		3		
34		2				1		
33		1				1		
32		0						
31		1						
N:	2,482	409	659	38	2,482	409	659	38
M:	50.43	42.85	47.20	42.95	50.76	42.59	47.45	42.47
SD:	3.51	3.70	3.29	2.84	3.52	3.53	2.93	2.96
M_1-M_2:	7.58		4.25		8.17		4.98	
t:	38.66		7.79		43.39		10.17	

male prison inmates. On Equation 1, the nondelinquents attained a mean of 47.20, standard deviation 3.29, and the delinquents a mean of 42.95, standard deviation 2.84. The difference of 4.25 gives rise to a *t* ratio of 7.79, which is significant beyond the .01 level. The statistics for the Italian delinquents are very similar to those observed for the 409 American delinquents in the cross-validating sample, but the nondelinquent Italian males have a somewhat lower average score on Equation 1 than their American counterparts.

On Equation 2 the Italian nondelinquents had a mean of 47.45, standard deviation 2.93, and the delinquents a mean of 42.47, standard deviation 2.96; the difference of 4.98 yielded a *t* ratio of 10.17, significant beyond the .01 level. Again, the figures for the Italian delinquents are very similar to those for the American delinquents, but the Italian nondelinquents scored lower than the American nondelinquents.

These significant *t* ratios for the two equations in the Italian comparisons give confirmation to the propositions advanced earlier about cross-cultural relevance and validity. The question now becomes one of screening efficiency. If the two equations are used to classify individuals as higher or lower on the social maturity continuum, what level of accuracy will be attained? The frequency distributions presented in Table 3 provide evidence on this query.

The degree of discrimination is to some extent visually apparent, and the four *t* ratios document the significance of these differences. A more precise indication of screening efficiency may be provided by the percentage of subjects in each distribution equaling or exceeding a specified cutting score. A convenient cutting score for the American samples would be 50. For Equation 1, 65% of the nondelinquents scored at or above 50, and 5% of the delinquents were in this category. For Equation 2, 66% of the nondelinquents

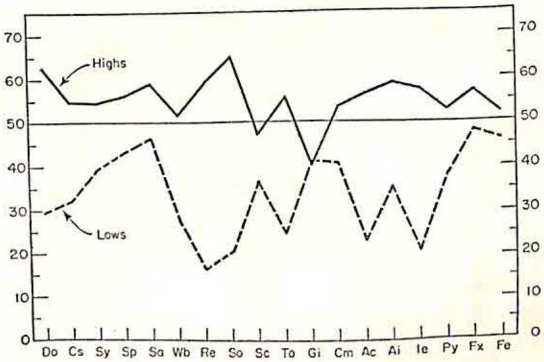


FIG. 1. Mean standard-score profiles on the CPI for the 25 highest scoring and 25 lowest scoring males on the 6-variable social maturity equation.

scored at or above 50, and 4% of the delinquents did the same.

A convenient cutting score for the Italian samples would be 47 (the mean on both equations). On Equation 1, 70% of the nondelinquents scored at 47 or above, versus 8% of the delinquents; on Equation 2, the percentages were 63 versus 5.

The ratio of nondelinquents to delinquents over all four comparisons is roughly 12–1, better than the 4–1 or 5–1 ratio which one might anticipate in a random ordering of cases.

Cutting scores could also be set so as to maximize the identification of delinquents. For scores of 44 and below, in the American samples, Equation 1 would identify 69% of the delinquents and 6% of the nondelinquents; the same cutting score on Equation 2 would identify 73% of the delinquents and 4% of the nondelinquents. For scores of 43 and below in the Italian samples, Equation 1 would identify 50% of the delinquents versus 8% of the nondelinquents, and for Equation 2 the figures would be 58% versus 8%.

To aid in visualizing the configurational relationships which are diagnostic of higher and lower degrees of social maturity, the two profiles in Figure 1 have been prepared.

From the cross-validating sample, the 25 men with highest scores on Equation 2 were identified, and a mean profile computed; this is the profile in Figure 1 drawn with a solid line. Similarly, a mean profile has been drawn (dotted line) for the 25 men with lowest scores on the equation.

On 17 of the scales, the more mature men score higher ($p \leq .05$ in each instance); on one scale, Good Impression, the more mature men score lower, although the difference is not significant. For the more mature sample, the highest points on the profile are on the scales for Dominance and Socialization, with next highest means on Responsibility, Self-Acceptance, and Achievement via Independence. Two scores are below the baseline of 50—Self-Control and Good Impression. The low points on the profile for the less mature subjects are on the scales for Responsibility and Socialization, with next lowest values on Intellectual Efficiency, Achievement via Conformance, and Tolerance.

The first profile is clearly indicative of an effective, well-socialized individual who is nonetheless free in the expression of impulse, and little bound by any superficial desire to please others or to do just what they want. The second profile, equally clearly, is that of an individual in severe friction with his environment, ineffective in utilizing his abilities, and dubious concerning the trustworthiness and integrity of others.

VALIDATION WITHIN NONDELINQUENT SAMPLES

A recent study by Hetherington and Feldman (1964) reported an analysis of college students who did or did not cheat on a set of course examinations.[3] By various techniques of observation and checking, it was discovered that 46 out of 78 students observed cheated one or more times. The two social maturity equations, if valid within the nondelinquent range, should differentiate between the two groups.

Equation 1 gave a mean of 50.78, standard deviation 4.46 for the 32 noncheaters, and 47.74, standard deviation 4.85, for the 46 who cheated. The difference of 3.04 was significant at the .01 level ($t = 2.78$). Because of the continuous nature of the distribution, it seems permissible to compute a biserial correlation; this was done, with a coefficient of +.39 being obtained. Using a cutting score of 51, 50 of the 78 students would be correctly classified as cheaters or noncheaters, an accuracy percentage of 64.

Equation 2 yielded a mean of 51.50, standard deviation 4.74, for noncheaters, and a mean of 48.72, standard deviation 4.87 for cheaters. The difference of 2.78 is significant at the .05 level ($t = 2.47$), and the biserial correlation coefficient was +.35. Again, using a cutting score of 51 or above, 47 of the 78 students would be correctly classified, an accuracy percentage of 60. The chance level of accuracy would be given by calling all students cheaters, which would give correct identification of 46, an accuracy percentage of 59.

The next data for validation come from a new sample of high school males ($N = 565$), not used in either the original or cross-validating analyses already reported. The principals in the schools where this testing was done were asked to nominate students who were especially high and especially low on the trait of social responsibility. Of the 565 students, 50 were nominated as "highs" and 51 as "lows" on responsibility. Arbitrary scores of 3, 2, and 1 were assigned to students in the high, unnominated, and low groups. Scores on Equations 1 and 2 were computed for all students, and correlated with the criterion of responsibility; the coefficients were +.38 for Equation 1 and +.37 for Equation 2. Both coefficients, although of only moderate magnitude, are highly significant statistically. The correlation between the two equations was +.95.

---

[3] Data from the original study and permission to conduct these new analyses were very kindly furnished by Hetherington and Feldman.

DESCRIPTIVE IMPLICATIONS

The final analyses to be reported concern the descriptive or interpersonal implications of the equations. The question asked is "what kind of person is it, in an everyday descriptive sense, who obtains a higher or lower score on the equations?" The instrument used to provide these descriptions is the Adjective Check List (Gough & Heilbrun, 1965), and the theoretical rationale underlying this method of specifying the meaning of test scores or other variables is presented in an earlier paper by the writer (Gough, 1965).

Three samples were utilized in the analyses: 100 military officers, assessed in a series of 5-day programs at the Institute of Personality Assessment and Research of the University of California; 41 members of a University of California fraternity, and 45 members of a University of California sorority. Staff psychologists provided the adjectival descriptions of the 100 officers, and the descriptions for the two living groups came from checklists completed by the members about each other. In each of the three samples, a descriptive score on each word of the List may be obtained by counting the number of times that word has been checked by the observers; these adjectival totals may then be correlated with any other variables which are to be studied.

For Equation 1, mean scores for the fraternity, sorority, and military samples, respectively, were 51.23, 54.13, and 50.27, with standard deviations of 2.95, 2.99, and 2.35. The 10 adjectives with highest positive correlations with Equation 1 in the military sample were: reserved, dependable, moderate, conscientious, steady, calm, painstaking, unassuming, wholesome, and reliable. The 10 words with lowest correlations were: temperamental, hurried, intolerant, disorderly, impulsive, frivolous, changeable, impatient, mischievous, and distractible. Within this sample of nondelinquent and relatively successful men, in other words, Equation 1 generates a distribution which has clear and distinct interpersonal implications.

Adjectival correlations with Equation 1 scores were computed separately for the college-student samples, and then averaged. The 10 highest correlations were: organized, reliable, responsible, capable, conscientious, mannerly, steady, self-controlled, tactful, and foresighted. The 10 lowest correlations were distractible, flirtatious, daring, reckless, unconventional, quick, defensive, careless, forgetful, and selfish. Among the college students, using peer descriptions as the source of observation, Equation 1 also generates a continuum having definite connotations for interpersonal behavior.

For Equation 2, the averages for the fraternity, sorority, and military samples were 52.18, 54.74, and 50.62, with standard deviations of 2.43, 2.88, and 2.43. For the military sample, the 10 words most differentially used to describe men with high scores were: rational, dependable, idealistic, reflective, calm, steady, wholesome, clear-thinking, organized, and reserved. The 10 words used most differentially to describe men with low scores were: shallow, intolerant, fickle, hurried, weak, nervous, temperamental, frivolous, zany, and distractible.

In the fraternity and sorority samples, the correlations for Equation 2 were computed separately, and then averaged as before. The 10 words used most typically to describe students with high scores were: tactful, reliable, capable, foresighted, mature, organized, dependable, responsible, stable, and wise. The 10 words used most typically to describe students with low scores were: defensive, unconventional, rude, coarse, distractible, noisy, obnoxious, blustery, loud, and tactless. The magnitude of these relations may be indicated by citing the averaged coefficients of +.39 for tactful, +.37 for reliable, −.35 for defensive, and −.30 for rude.

DISCUSSION

Social maturity, or interpersonal maturity, is a behavioral criterion which ought to be predictable from the CPI. Furthermore, it is a criterion having cross-cultural status, and which therefore should be predictable from the same basic pattern of CPI variables in any culture.

The first step in the analysis of this problem is to define a pattern or cluster of variables on the inventory which can identify individuals manifesting higher or lower levels of maturity. Two such equations were developed, one containing only three variables and adapted for hand calculation, and the other containing six scales and recommended for use where machine facilities are available and where precise specification of the maturity continuum is desired.

Both equations appear to be quite effective in distinguishing between more and less delinquent samples in the United States, and the differentiation was almost as good in an Italian comparison. Correlations with other indices of interpersonal maturity, such as cheating versus not cheating in course examinations and ratings of social responsibility, were statistically significant although not high.

Perhaps the strongest evidence pertaining to the psychological meaning of the equations comes from the adjectival analyses. In three samples, including males and females, students and adults, and peer ratings and ratings by psychologists, there was clear specification of a continuum of interpersonal implications. High scorers on both equations are consistently described as dependable, reliable, responsible, mature, and capable, and low scorers are consistently seen as distractible, impulsive, defensive, temperamental, and intolerant.

The variables included in the equations are consonant with these findings, and with the theoretical goal of assessing the construct of social maturity. The trait combination, as delineated in Equation 2, stresses socialization and responsibility, to be sure, but also includes emphases on flexibility and constructive leadership. And at the other end it weights negatively any disposition to respond in a purely favorable or perfunctory manner. The high-scoring man, one could say, although basically stable and self-disciplined is nonetheless open to new experience, relatively free of petty conventionality and the suppression of impulse, and hence in a position to alter as well as to preserve his culture.

REFERENCES

GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

GOUGH, H. G. Theory and measurement of socialization. *Journal of Consulting Psychology,* 1960, **24**, 23–30.

GOUGH, H. G. Academic achievement in high school as predicted from the California Psychological Inventory. *Journal of Educational Psychology,* 1964, **55**, 174–180.

GOUGH, H. G. The conceptual analysis of psychological test scores and other variables. *Journal of Abnormal Psychology,* 1965, **70**, 294–302.

GOUGH, H. G., & HEILBRUN, A. B., JR. *The Adjective Check List Manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1965.

GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology,* 1952, **16**, 207–212.

HETHERINGTON, E. M., & FELDMAN, S. E. College cheating as a function of subject and situational variables. *Journal of Educational Psychology,* 1964, **55**, 212–218.

MACKINNON, D. W., CRUTCHFIELD, R. S., BARRON, F., BLOCK, J., GOUGH, H. G., & HARRIS, R. E. An assessment study of Air Force officers. Part I: Design of the study and description of the variables. *USAF WADC Tech. Rep.,* 1958, No. AD 151 040.

(Received May 6, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 196–208

RESPONSIVENESS TO SUGGESTIONS FOLLOWING WAKING AND IMAGINATION INSTRUCTIONS AND FOLLOWING INDUCTION OF HYPNOSIS [1]

ERNEST R. HILGARD AND CHARLES T. TART [2]

Stanford University

2 experiments are reported to test the increase of responsiveness to suggestion tests following hypnotic induction over responsiveness to such tests in waking and imagination conditions, an increase that has been doubted as a result of experiments by Barber and Calverley (1962, 1963). In the 1st experiment, 60 Ss were divided into groups of 20 Ss each serving under 1 of 3 conditions in a 1st session (waking, imagination, hypnosis). All received a standard hypnotic induction in a 2nd session. While the treatment effects did not yield significant differences on the 1st day, there were significant gains in responsiveness to suggestions by the waking and imagination groups in the 2nd session. In the 2nd experiment, with some methodological improvements, 90 Ss served in 6 groups of 15 Ss each, in imagination without expectation of hypnosis, imagination with expectation of hypnosis, and hypnotic induction, in various combinations. Significant gains were found with hypnotic induction throughout. State reports (subjective responses of drifting into hypnosis) showed that those Ss within both imagination and hypnotic induction conditions who reported themselves as becoming hypnotized were the ones who yielded the highest suggestibility scores. The difficulty of obtaining significant treatment effects is noted unless Ss serve as their own controls.

Increase in suggestibility has so long been associated with hypnosis that such increase has commonly been considered to be a defining characteristic of hypnosis. Thus the titles of standard books reflect this intimacy between hypnosis and suggestibility, for example, *Hypnosis and suggestibility* (Hull, 1933), *Hypnotism: An objective study in suggestibility* (Weitzenhoffer, 1953). It is well-known, however, that responses to suggestions of the kind given within hypnosis may be obtained outside the hypnotic state; such suggestions are often called waking suggestions, a term of convenience that can be used without implying that hypnosis is a sleep state. When comparisons are made between responsiveness to suggestions in the waking state and following hypnotic induction, two facts stand out:

1. There is a substantial positive correlation between responsiveness within the waking state and following hypnotic induction. This implies that responsiveness to suggestions carries within it a heavy loading of individual differences common to both the waking and hypnotic states.

2. The increase in responsiveness to suggestions over the waking state engendered by a prior hypnotic induction is usually found to be significant, but the gain is typically small "probably far less than the classical hypnotists would have supposed had the question ever occurred to them [Hull, 1933, p. 298]." Hull's conclusions of a significant but small gain following hypnotic induction have been verified by Weitzenhoffer and Sjoberg (1961) and by Barber and Glass (1962).

The second of these summary conclusions has been challenged by Barber and Calverley (1962, 1963) on the grounds that the gains in suggestibility found experimentally are to be attributed to circumstances of experimentation, such as aroused task motivation, rather than to the hypnotic state as such. When they aroused task motivation in both the waking state and the hypnotic state, differences between the two states became minimized or disappeared altogether. Their conclusions rest upon an assumed equation of the task motivations aroused under the two conditions. Unfortunately, their instructions differed in important respects between their waking groups and their hypnotic groups. In addition, by contrast with the earlier Barber and Glass (1962) experiment, they no longer used subjects as their own controls, so that other differences, besides the added task motivation, occur between the experiments.

---

[1] The investigation was carried out in the Laboratory of Hypnosis Research, Department of Psychology, Stanford University, with the aid of a grant from the National Institute of Mental Health, Public Health Service (Grant MH-03859). Grateful acknowledgement is made of the assistance of the staff and research assistants in the Laboratory for their aid in the conduct of the experiments reported. At the time C. T. Tart was holding a Public Health Service Postdoctoral Research Fellowship (1-F2-MH-14,622-01).

[2] Now at the University of Virginia.

There are important methodological considerations that arise when two experimental conditions are to be compared, when (*a*) responses between the two sets of conditions are highly correlated, and (*b*) when the expected differences are small. If the preferred hypothesis calls for no difference, this result is more likely to emerge if group comparisons are made, and the correlation term neglected, as in the Barber and Calverley studies. If one seeks to detect any difference, however small, then the subject is likely to be used as his own control, to take advantage of the correlational term. The considerations are not entirely statistical ones, however, but depend on the appropriateness of what is done to the actual relationships being studied. Because these problems are peculiarly difficult within hypnosis, some additional comments are in order before proceeding to a description of the current experiments.

There are good reasons for not using subjects as their own controls within hypnotic experiments, particularly because of tendencies, whether deliberate or unconscious, for subjects to modify their behavior under control conditions when they know that a hypnotic session is to follow (Zamansky, Scharf, & Brightbill, 1964). There is no assurance, however, that expectations are circumvented by having only one condition per subject either, if a subject knows that other subjects are being hypnotized. Failure to use subjects as their own controls greatly attenuates the possibility of finding differences under two conditions of experimentation. This follows because:

1. The variability in responses to hypnotic suggestions from one subject to another is very great, the distributions are often flat, sometimes bimodal, with extreme scores at both ends of the scale (Hilgard, Weitzenhoffer, Landes, & Moore, 1961). Under these circumstances two risks are taken in comparisons between small groups: First, that genuine differences may be obscured because they are small and the variability great; second, that obtained differences, meeting conventional statistical standards of significance, may arise when assignment is random through the chance appearance in small groups of disproportionate numbers of extremely high-scoring or extremely low-scoring subjects.

2. The nature of responsiveness to hypnosis is such that within a random sample there will appear a substantial number of essentially nonsusceptible subjects, perhaps as many as two thirds of the group. Such subjects will dilute any comparison between randomly selected groups assigned to waking and hypnotic conditions. Because they are relatively insusceptible, gains between waking and hypnosis are not to be expected; because they produce for the most part extremely low scores they will contribute to low means and high variability relative to the mean, thus reducing the probability of significant differences between the compared groups.

3. At the other end of the scale, a few subjects with very high scores enter hypnosis spontaneously without a prior hypnotic induction, and for them also the changes from waking to hypnosis will be slight or absent. Because their scores are high, they also increase the variability that reduces the significance of differences between means.

As a consequence, in group comparisons, the weight must be carried by those few subjects who make substantial gains between waking and hypnosis. An overall comparison between two groups that does not take into account the correlation between waking responsiveness and hypnotic responsiveness will understate the gains that take place for some subjects.

In view of these considerations, the experiments to be reported have been carried out in such a manner as to reveal some of the finer grain of the changes that take place between waking and hypnotic conditions, and the consequences of alternative treatments of the data. To this end, the instructions followed are essentially those of Barber and Glass (1962), rather than the task-motivation instructions of Barber and Calverley (1962, 1963). It is conjectured that the Barber and Calverley results do not require the task-motivation interpretation at all; that is, had the Barber and Glass subjects not served as their own controls, significant changes would not have been found. The design of the experiments to be reported permits both types of comparison, that is, one in which random groups are compared, and one in which the subject serves as his own control.

EXPERIMENT I. WAKING INSTRUCTIONS, IMAGINATION INSTRUCTIONS, AND HYPNOTIC INDUCTION

Instructions and procedures. The instructions in this experiment were modeled very closely after those of Barber and Glass (1962). Of the 60 subjects, from an introductory psychology class, 20 were randomly assigned to each of three treatment conditions on the first day of the experiment, the groups being designated Waking, Imagination, and Hypnotic Induction. On the second day, all subjects received the hypnotic induction.[3] Thus the first day permitted comparisons among randomly assigned subjects (following the Barber and Calverley data treatment), while the second day permitted treating the subjects as their own controls (following the Barber and Glass treatment).

For the waking condition, the following statement was made to the subject prior to his receiving the test suggestions:

> Today you will be given tests of responsiveness to waking suggestions. You will not be hypnotized. We want to be sure that you do not become hypnotized; if you slip into a hypnotic state inadvertently, as indicated by your state report, we will bring you back to your normal state. It is still possible to respond well to hypnotic-like suggestions in the wide-awake normal state.

For the imagination condition, the following statement, adapted from Barber and Glass (1962), was read to the subject:

> Today you will be given tests of imagination under normal conditions. You will not be hypnotized. The better you can imagine, the more you'll respond; try as hard as you can to concentrate and to imagine that the things I tell you are true.

It may be noted that a deliberate effort was made to keep the waking subjects from drifting into hypnosis; while the imagination subjects were told that they would not be hypnotized, nothing was said about spontaneous hypnosis.

In order to control the drift into hypnosis in the waking condition, and to assess the extent to which the subject became spontaneously hypnotized under the imagination condition, or following suggestions within the hypnotic induction, a "state" report was frequently called for.

The "state" report represented a modified version of such reports earlier used by LeCron (1953), Hatfield (1961), Tart (1963). Before experimentation began the subject was taught how to use the scale, four numerals from 0 ("zero") to 3. He was told to respond "zero" to the question "State?" if he felt in his usual, normal, wide-awake state, to respond "1" if he felt himself very relaxed or drifting off, as in going to sleep, to respond "2" if he felt himself in a mildly hypnotic state, and "3" if more deeply hypnotized. Because depth of hypnosis often fluctuates within a session, these reports were called for repeatedly within the session, typically before and after each test suggestion. If a subject in the waking condition gave a report of "1" he was aroused so that he returned to a report of "0."

For the hypnotic condition, the standard induction method of Form C of the Stanford Hypnotic Susceptibility Scale (SHSS) was followed (Weitzenhoffer & Hilgard, 1962). This induction involves eye fixation and eye closure, along with suggestions of relaxation and sleep, and requires some 10 minutes. It may be noted that the waking and imagination instructions were much briefer. Exactly the same set of test suggestions followed waking, imagination, and hypnotic instructions. These represented 10 items of SHSS, Form C, omitting the taste hallucination and the hallucinated voice, in order to shorten the time required. An alternate but equivalent form, designed for the purpose, was used on the second day.[4]

Results of Experiment I. The major results are shown in Table 1. While the means of the first day are in ascending order by conditions, a simple analysis of variance, based on the one day only, shows that, as anticipated, they do not differ significantly. If these were the only data collected, the authors would agree with the conclusions of the Barber and Calverley studies that hypnotic induction adds nothing significant to suggestibility,

---

[3] For another purpose to supply a pool of subjects contrasted in their waking and hypnotic responsiveness, another 20 subjects were run at the same time under the waking-hypnotic condition. The 20 subjects treated as paralleling the imagination and hypnotic condition represent a random half of the total 40 subjects run.

[4] A limited supply of the alternate forms is available on request to the Laboratory of Hypnosis Research, 582 Alvarado Row, Stanford, California.

TABLE 1

RESPONSES TO SUGGESTIONS IN WAKING CONDITION, IMAGINATION CONDITION, AND FOLLOWING HYPNOTIC INDUCTION: EXPERIMENT I

	Scores on suggestion tests (maximum = 10)		Difference: change from first to second session
	First session	Second session	
Group 1 ($N = 20$)	Waking condition *M* 2.88 *SD* 2.78	Hypnotic induction *M* 4.90 *SD* 2.41	*M* +2.02 (t test, $p < .0005$)
Group 2 ($N = 20$)	Imagination condition *M* 3.70 *SD* 3.15	Hypnotic induction *M* 5.70 *SD* 2.79	*M* +2.00 (t test, $p < .005$)
Group 3 ($N = 20$)	Hypnotic induction *M* 4.90 *SD* 2.50	Hypnotic induction *M* 4.55 *SD* 2.53	*M* −0.35 (t test, *ns*)

Note.—Results of overall analysis of variance: Group effects, *ns*; first session differs from second, $p < .001$; Interaction, Groups × Sessions, $p < .01$. Results of simple analyses: Group effects, First session, *ns*. Group effects, Second session, *ns*.

even though no special task motivation is added, within the waking and imagination conditions. If the assignment of subjects was truly random, and there were no order-of-session effects, no differences would be expected on the second day either. A simple analysis of variance on this day also confirmed the conjecture of nonsignificant differences. There were, however, significant gains between the two days for both the waking and imagination groups, and no gain (a trivial loss instead) for the group experiencing hypnotic induction on both days, a result that Barber and Glass (1962) had also found.

The reason that it is possible to detect the effects of induction in this design, and not in the random assignment one alone, is that this design takes into account the correlations between waking and hypnosis scores. In this experiment these were: for waking-hypnosis, $r = .65$; for imagination-hypnosis, $r = .66$; for hypnosis-hypnosis, $r = .87$.

Although the essential information is given in the simple analyses of the data as reported, it is possible to make a somewhat more elegant analysis through the use of an over-all analysis of variance, in order to find an interaction term. By the use of Case 15 (McNemar, 1962), the results noted at the bottom of the table are arrived at, namely, that there is no significant overall treatment effect, but there is a significant gain overall with hypnotic induction on Day 2 ($p < .001$) and there is a significant interaction effect ($p < .01$) between the treatment conditions and the days, which the t tests pinpoint.

The experiment just described is generally satisfactory from two points of view: (*a*) by using their design for group comparisons, it replicates the nonsignificant findings of Barber and Calverley (1962, 1963), without requiring an explanation in terms of task motivation, and (*b*) by adding a session to permit the study of changes by individual subjects, significant differences are found, confirming the findings of Barber and Glass (1962), and indicating that the Barber and Calverley design is not sensitive to small but significant changes that may indeed occur within hypnosis.

There is a flaw in the current experiment, however, which is also in common with the experiments replicated, in that the instructions were given verbally by a hypnotist who conceivably might have unwittingly biased the results. The experimenters were of course aware of this danger, and sought to surmount it by reading all instructions verbatim and by changing experimenters from one day to the next, the experimenter lacking knowledge of the subject's performance on the prior day. A control experiment showed, however, that the experimenter's voice was demonstrably different after he had spent 10 minutes in inducing hypnosis than after he had merely read off very short waking or imagination instructions (Troffer & Tart, 1964). While this may not have made any difference, a second experiment was designed with tighter controls, in which this flaw was eliminated.

This second experiment will be described before attention is directed to the state reports from both experiments.

EXPERIMENT II. IMAGINATION INSTRUCTIONS WITH AND WITHOUT EXPECTATION OF HYPNOSIS AND HYPNOTIC INDUCTION

Instructions and procedures. While the main new control in this experiment was the use of taped instructions and suggestion tests, with the same master tape throughout, to circumvent any possible biasing hypnotist effect, several other changes were made in order to obtain additional information about the details of what goes on within a waking-imagination session as compared with a usual hypnotic one. Three different instruction conditions were used in various combinations. These included, first, an imagination condition without expectation of hypnosis (hereafter designated INE, for Imagination–No-Expectation), second, an imagination condition in which the expectation was aroused that this would lead to a hypnotic state (designated IE, for Imagination with Expectation), and, third, the usual standard hypnotic induction of SHSS, Form C (designated HYP). The uninstructed waking condition of Experiment I was dropped, and no effort was made, beyond the initial instructions, to prevent a spontaneous drift into hypnosis. State reports were called for as before, the voice on the tape requesting them at the beginning and end of each suggestion.

Another change was introduced, in that the time allowed for the subject "to exercise his imagination" was equal to the time spent in hypnotic induction, so that, if relaxed immobility for 10 minutes or so is important, it was equally there for the "waking" conditions and for the "hypnotic" ones.

The exact wording of the instructions for those who were in the Imagination-No-Expectation of Hypnosis condition (INE) was as follows:

> Today you will be given tests of imagination while you are relaxed. It has been generally found, however, that exercising your imagination strongly in these experimental tasks will allow you to experience the things I say quite vividly, even though you have not been hypnotized. The better you imagine, the more you'll respond. Try as hard as you can to concentrate, and to imagine that the things I tell you are true.

In addition, the subjects were told whether or not they would receive a formal induction on the second day. Separate tapes were prepared for each condition, but where the instructions were uniform, a duplicate of the original master tape was used, so that there would not be variations in the voice, even though the same voice (that of E.R.H.) was used throughout. An assisting experimenter (not E.R.H.) was always present, and the tape indicated when responses should be made so that this experimenter could record the response. The role of the assisting experimenter was minimal, conducting the subject to the room, occasionally holding a bottle under his nose, or preparing something for him to look at, turning on the tape recorder, keeping score. The experimenter never questioned the subject, except for an interrogatory at the very end of the second day.

The corresponding wording of instructions for those in the condition calling for Imagination-Expectation of Hypnosis (IE) was:

> Today you will be given tests of imagination while you are relaxed. It has generally been found that exercising your imagination strongly in these experimental tasks produces a hypnotic state, even though we don't go through the formalities of inducing hypnosis. The better you imagine, the more you'll respond. Try as hard as you can to concentrate and to imagine that the things I tell you are true.

These subjects were also informed whether or not they would receive a formal hypnotic induction on the second day.

Those who were to receive the formal hypnotic induction (condition HYP) were told:

> Today you will be hypnotized and given a number of hypnotic tests. It has generally been found that exercising your imagination during the hypnotic induction and throughout the whole experiment today will induce and maintain a hypnotic state. The better you imagine, the more you'll respond. Try as hard as you can to concentrate and to imagine that the things I tell you are true.

These instructions were then followed by the standard induction of SHSS, Form C.

In order to control for any of the "holding back" that is attributed to hypnotic subjects when they know that a hypnotic session is to follow a control one, the control subjects, who had nonhypnotic tests on both days, were told clearly that they were in control groups and would not be hypnotized on either day. No deception was practiced; no subject who was told he was not to be hypnotized was hypnotized on either session. Those who were not hypnotized on the first day, but were hypnotized on the second, were told that they would be hypnotized on the second day. By making full disclosure the subject's "guessing" what would happen was avoided. It was thus possible to compare for each condition what happened when there was expectation of either uniformity or change of conditions on the next day. The total design can be summarized most simply by a table describing the groups and treatment conditions.

The design can most simply be understood as three subexperiments in which pairs of groups have identical conditions on the first session. For one member of the pair the condition is repeated on the second day as a control for practice effects and expectations; the other group has another condition on the second day. Other pairings of conditions could of course be made, but these seemed to be

TABLE 2

DESIGN OF EXPERIMENT II

Group	First session	Second session
1. INE-INE ($N = 15$)	Imagination, no expectation of hypnosis	Imagination, no expectation of hypnosis
2. INE-HYP ($N = 15$)	Imagination, no expectation of hypnosis	Hypnotic induction
3. IE-IE ($N = 15$)	Imagination, expectation of hypnosis	Imagination, expectation of hypnosis
4. IE-HYP ($N = 15$)	Imagination, expectation of hypnosis	Hypnotic induction
5. HYP-HYP ($N = 15$)	Hypnotic induction	Hypnotic induction
6. HYP-INE ($N = 15$)	Hypnotic induction	Imagination, no expectation of hypnosis

the most instructive. There were 15 subjects in each condition, 90 in all, from classes in introductory psychology.

When the experiment was entirely completed, many of the subjects who had served in the control groups expressed disappointment at not having had the experience of hypnosis, and requested return to the laboratory. This provided the circumstances to invite back for a third day all of the subjects who had been in the control groups, offering a new opportunity for hypnosis without violating the agreements made with them during the experiment proper. At the same time, the opportunity was taken to invite back also those who had had hypnosis only, in order to test them under imagination conditions. Responses to these postexperimental invitations were accepted by 18 of the 30 subjects who had been in imagination groups on both days, and by 6 of the 15 subjects who had been in hypnosis groups both days. Their results will be considered after the results of the experiment proper have been reported.

Results of Experiment II. The first task is to inspect the results as they would be were the data from the first day only to be considered, with 30 subjects in each of the three conditions. This is not a favorable type of comparison, for the reasons indicated earlier, and will detect differences only if they are more substantial than those found in Experiment I. The results are presented in Table 3.

Contrary to expectations, there is a significant treatment effect, an analysis of variance showing that group effects are significant between $p = .01$ and $.025$. The effects are not striking, the Duncan multiple-range test showing no difference between neighboring groups, but a significant difference between the extremes. It should be noted also that the sizes of these groups have been increased to 30 subjects each. Barber and Calverley (1963) also found some differences of this kind; when groups are large enough, and differences great enough, a less sensitive method can of course detect them.

Returning now to an analysis in which the subjects serve as their own controls one expects to find more significant differences, although many of the comparisons will be made with groups of 15 subjects instead of 30 as in Table 3.

First to be considered is a control comparison, in which no differences are to be expected. These are the results for those groups that had the same conditions on each of two days. If there are differences they will represent practice effects, or perhaps slight differences between the two forms of the test. The results are given in Table 4. The analysis of variance shows all effects to be nonsignificant. These results permit acceptance with greater confidence of the significant changes to be attributed in subsequent analyses to treatment effects. They also establish the comparability of the suggestibility forms used on the 2 days. The unusually low mean hypnotic scores of Group 3′ illustrate what was said earlier about chance effects with small groups; this group of 15 subjects happened to lack any highly susceptible subjects. The low scores are not attributed either to the use of taped inductions or to expectations owing to having hypnotic induction on

TABLE 3

RESPONSIVENESS TO SUGGESTIONS WITHIN THE FIRST SESSION UNDER TWO IMAGINATION CONDITIONS, AND FOLLOWING HYPNOTIC INDUCTION: EXPERIMENT II

Condition	Scores on suggestion test (Modified Form C, maximum = 10)
Imagination, no expectation of hypnosis (INE)	$N = 30$ $M = 2.58$ $SD = 2.27$
Imagination, expectation of hypnosis (IE)	$N = 30$ $M = 3.37$ $SD = 2.12$
Hypnotic induction (HYP)	$N = 30$ $M = 4.43$ $SD = 2.75$

Note.—Results of analysis of variance: For Groups, $F = 4.52$, $.025 < p < .01$.

TABLE 4

CONTROL COMPARISONS: EFFECT OF REPETITION OF SAME CONDITION UNDER TWO IMAGINATION CONDITIONS AND FOLLOWING HYPNOTIC INDUCTION: EXPERIMENT II

	Scores on suggestion tests (Maximum = 10)		Difference: change from first to second session
	First session	Second session	
Group 1′ ($N = 15$)	Imagination, no expectation of hypnosis *M* 2.37 *SD* 1.99	Imagination, no expectation of hypnosis *M* 2.33 *SD* 1.71	*M* −.04 (*t* test, *ns*)
Group 2′ ($N = 15$)	Imagination, expectation of hypnosis *M* 2.73 *SD* 1.99	Imagination, expectation of hypnosis *M* 3.23 *SD* 1.81	*M* +.50 (*t* test, *ns*)
Group 3′ ($N = 15$)	Hypnotic induction *M* 3.83 *SD* 2.91	Hypnotic induction *M* 3.77 *SD* 2.80	*M* −.06 (*t* test, *ns*)

Note.—Results of analysis of variance: Groups effect, *ns*; session effects, *ns*; interaction, Groups × Sessions, *ns*. Two-day correlations: Group 1′ (INE versus INE), $r = .65$ ($N = 15$); Group 2′ (IE versus IE), $r = .40$ ($N = 15$); Group 3′ (HYP versus HYP), $r = .98$ ($N = 15$).

both days. Expectations are no different from those subjects in Group 3 of Experiment I (Table 1), who averaged a point higher. The taped induction can scarcely be blamed because the other 15 subjects of Experiment II who were hypnotized by tape on the first day averaged 5.03 compared with the first day average of 4.90 for the 20 subjects of Experiment I, in which induction was live. A random effect in the composition of a small sample of subjects remains the best explanation of the low mean hypnotic score of Group 3′.

The main results are found in the relationships of Table 5. The analysis of variance shows the interaction effects of groups and sessions to be highly significant ($p < .005$), which means that the gains from imagination conditions to hypnotic induction (and the loss when hypnotic induction comes first) have been demonstrated. The *t* tests for each group are satisfactory, particularly

TABLE 5

RESPONSES TO SUGGESTION IN TWO IMAGINATION CONDITIONS AND FOLLOWING HYPNOTIC INDUCTION: EXPERIMENT II

	Scores on suggestion test (Maximum = 10)		Difference: change from first to second session
	First session	Second session	
Group 1 ($N = 15$)	Imagination, no expectation of hypnosis *M* 2.80 *SD* 2.57	Hypnotic induction *M* 3.70 *SD* 3.08	*M* +.90 (*t* test, $p < .025$)
Group 2 ($N = 15$)	Imagination, expectation of hypnosis *M* 4.00 *SD* 2.11	Hypnotic induction *M* 5.13 *SD* 2.42	*M* +1.13 (*t* test, $p < .025$)
Group 3 ($N = 15$)	Hypnotic induction *M* 5.03 *SD* 2.52	Imagination, no expectation of hypnosis *M* 3.53 *SD* 2.38	*M* −1.50 (*t* test, $p < .025$)

Note.—Results of analysis of variance: groups, *ns*; sessions, *ns*; interaction, Groups × Sessions, $p < .005$; 2-day correlations: Group 1 (INE versus HYP), $r = .86$ ($N = 15$); Group 2 (IE versus HYP), $r = .72$ ($N = 15$); Group 3 (HYP versus INE), $r = .52$ ($N = 15$).

TABLE 6

CHANGES IN RESPONSIVENESS WITHIN A THIRD SESSION ORIGINALLY UNPLANNED

	Scores on suggestion test (Maximum = 10)		Difference: change from second to third session
	Second session (repeating conditions of first)	Third session (unplanned)	
INE-INE-HYP (N = 8)	Imagination, no expectation of hypnosis M 2.75 SD 0.60	Hypnotic induction M 4.00 SD 1.73	M +1.25
IE-IE-HYP (N = 10)	Imagination, expectation of hypnosis M 3.55 SD 1.83	Hypnotic induction M 3.10 SD 2.20	M −0.45
HYP-HYP-INE (N = 6)	Hypnotic induction M 5.00 SD 3.26	Imagination, no expectation of hypnosis M 2.42 SD 1.71	M −2.58

Note.—t test, all hypnotic induction scores minus all imagination scores, $p < .05$.

in view of the small numbers of subjects in each group. Within each day the treatment means are not significantly different, reflecting the smaller sample size as compared with the significant first-day effects presented in Table 3.

The additional data from the subjects who returned for a third day suggest another comparison: What happens when an unanticipated third session provides a change from the uniformly experienced 2 days of the regular experiment? Although the groups are small, the overall results are significantly in the same direction as the main findings (Table 6). The only reversal is a failure of those who had had 2 days of imagination with expectation of hypnosis to improve under hypnotic induction, but this result, with 10 subjects only, is too slight to destroy the mean effect for the total of 24 subjects participating in this third session.

The results of the second experiment confirm those of the first: In all comparisons, with the minor exception just noted, hypnotic induction produces an increase in response over noninduction conditions, a gain which, in its overall effect, meets acceptable statistical standards of significance.

INFORMATION ADDED BY THE STATE REPORTS

In view of the large number of subjects insusceptible to hypnotic induction it would be very desirable, in experiments of this sort, to be able to compare results only for subjects capable of influence by hypnotic induction. The state reports are of help in this respect, because they are relatively independent of the responses to the various tests of suggestion. An attempt will be made to show that they do indeed provide information that is supplementary, and throw additional light upon what is happening in experiments of this kind.

Reliability of state reports. Within any 1 day there were a number of state reports obtained, so that a mean state report could be computed for each subject with a value between 0 and 3. The reliabilities of such reports are shown by the 2-day correlations of Table 7, in which all groups with the same condition on 2 days have been studied. For the hypnotized groups, these are in the .90's; while for the nonhypnotized groups they are lower, they are still in the .80's. Any spread in scores for the nonhypnotized

TABLE 7

RELIABILITY OF STATE REPORTS

	Common treatment on 2 days	Correlation between mean state reports
Experiment 1	HYP-HYP	.91 (N = 20)
Experiment 2	HYP-HYP	.99 (N = 15)
	IE-IE	.83 (N = 15)
	INE-INE	.81 (N = 15)

TABLE 8

CORRELATION BETWEEN STATE REPORTS AND OBJECTIVE SUGGESTION SCORES, ALL CONDITIONS

Subject group	Correlation between mean state report and suggestibility scores
Experiment I	
Waking (N = 40) (first day)	.22[a]
Imagination (N = 20) (first day)	.67
Hypnosis (N = 80) (second day only)	.68
Experiment II	
Imagination, no expectation of hypnosis (INE) (N = 45) (first day with this condition)	.65
Imagination, expectation of hypnosis (IE) (N = 15) (first day only)	−.01
Hypnosis (N = 60) (first day with this condition)	.75

[a] This is attenuated by the fact that the subject was aroused if a state report above 0 was given. Even so, enough subjects yielded such state reports often enough to produce the small correlation.

groups means, of course, that some have drifted into a borderline state or one that they were willing to characterize as similar to hypnosis; otherwise there would be no variance and no correlation.

State score a function of instructions and inductions, not of responsiveness to suggestibility to test items. There is usually a correlation between the state report and responsiveness to suggestions, as shown in Table 8. The one failure of correlation, that for the IE group, is puzzling, but the number of cases is small ($N = 15$) and it happens that two subjects giving high state reports were very low in their suggestibility scores. Such deviant subjects can destroy the correlation in a small sample of this kind. The other positive correlations may be taken as much more typical, based as they are on large samples. The positive correlations between state reports and responsiveness to suggestions raise a ticklish problem. May it not be that the state reports are themselves reflections of the fact that the subject finds himself responding to suggestions? The answer cannot be a straightforward one, for suggestions of drowsiness, and so on, if responded to, in part characterize the "state" that is being reported. But there are some data that show that this is not simply an alternative form of report on suggestibility, based on the subject's self-scoring of his responses within the suggestibility tests.

TABLE 9

PREDICTABILITY OF LATER STATE REPORTS FROM INITIAL REPORT FOLLOWING INSTRUCTIONS OR HYPNOTIC INDUCTION: EXPERIMENTS I AND II COMBINED, FIRST DAY ONLY

Initial state report	Mean state report within day 1 (disregarding condition)			
	0–0.9	1.0–1.9	2.0–2.9	Total
0	92	5	0	97
1	15	29	6	50
2–3	0	6	17	23
	107	40	23	170

Note.—$\chi^2 = 153.89$, $p < .001$, 4 df. $C = .69$ (contingency coefficient).

The first point to be made is that the initial state report, following waking instructions, imagination instructions, and hypnotic induction, tends to persist throughout the session, *despite the subject's later responsiveness or lack of responsiveness to the suggestions.* The facts for the first day of experimentation are summarized in Table 9, which combines the results for both experiments. It is evident that the initial report is highly predictive of later ones: $\chi^2 = 153.89$ with 4 df, $p < .001$, and the contingency coefficient (Siegel, 1956) is .69. Of those who start with but a report of 0 (normal aroused waking state), 92 of 97 average below 1.0 in their later reports, and none averages as high as 2.0. Correspondingly, of those who initially report 2 or 3 (hypnotized) 17 of 23 remain there, and none averages below 1.0. Thus, despite the correlation between the state report and the later suggestibility measures, the state report does not appear to depend upon the self-observation of these later experiences.

The second point is that the statistical regression may differ by conditions, even when the correlations between state report and suggestibility scores are nearly alike. Useful comparisons from Experiment I are available, in which the state reports correlate with suggestibility .67 for the imagination group and .68 for the hypnosis group, and another comparison from Experiment II, in which the correlations are .65 for the INE group and .75 for the HYP group. When the regression lines are plotted in raw score form, it becomes evident that for an equal amount of suggestibility response there is a lesser state report for the nonhypnotized than for the hypnotized subjects (Figure 1). This argues also for some separateness between the judgments of state and those of responsiveness to suggestions.

The state reports furnish a somewhat crude but useful answer to the need for evidence that a person is hypnotized, apart from responsiveness on tests of suggestibility, for, as just indicated, there is no absolute correspondence between hypnotic state and responsiveness to standardized suggestions. If one were willing to accept the state measure following either waking instructions or hypnotic induction as a sorting measure, eliminating those from the waking group who feel hypnotized despite lack of induction, and from the hypnotic group those who do not feel hypnotized despite the attempted induction, the obtained differences between waking and hypnosis would be greatly enhanced. This is a tricky matter, of course, because of the extent to which the state report is itself a response to suggestions. Yet some such sorting basis would be highly desirable. An illustration of the difference that it would make is presented in Table 10 which shows the suggestibility scores following imagination instructions and following hypnotic induction of those who gave various state reports at the end of the instruction or induction period but before the tests of suggestion were made. It can be seen that those who felt themselves unaffected by the instructions or hypnotic induction (state report 0) scored at the bottom of the group, *whether or not there had been attempted hypnotic induction*. On the other hand, those five subjects who gave state reports of 2 ("hypnotized") following imagination instructions yielded suggestibility scores equaling those of the subjects who gave the same report following a standard hypnotic induction. Thus in group comparisons in which various alternatives to induction are used, it may be presumed that some subjects actually become hypnotized. The similar means for responsiveness to suggestions after equivalent initial state reports are striking; of course the numbers of cases in each category reflect the greater effectiveness of the hypnotic induction.

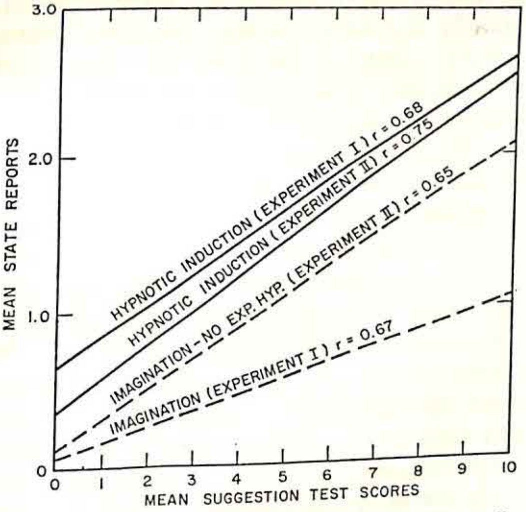


FIG. 1. Regression lines for state reports according to degree of responsiveness to suggestions. (Note that despite similar correlations, higher suggestibility scores are associated with lower state reports under imagination conditions than under hypnotic conditions.)

TABLE 10

SUGGESTIBILITY SCORES FOLLOWING VARIOUS INITIAL STATE REPORTS AFTER EITHER IMAGINATION INSTRUCTIONS OR ATTEMPTED HYPNOTIC INDUCTION: EXPERIMENT II

Initial state report, prior to suggestibility tests	Suggestibility scores (Modified Form C)					
	Imagination instructions (Either INE or IE)			Hypnotic induction (HYP)		
	N	*M*	*SD*	*N*	*M*	*SD*
3–very hypnotized	0	—	—	5	6.9	1.9
2–hypnotized	5	5.0	2.2	21	5.7	2.5
1–drifting	38	3.5	2.2	43	4.0	2.2
0–normally awake	38	2.2	0.8	12	1.3	1.4
Total[a]	81	3.0	1.5	81	4.2	2.1

[a] The groups of 81 subjects have 63 subjects who served in both conditions. If a subject served in the same condition twice, the first experience only was used.

The relationship between waking scores and hypnotic scores for individuals. Information about individuals is easily lost within pooled data. For example, the relatively high correlations between waking and hypnotic performances indicate that they have something in common, but they do not reveal anything about individual departures from this generalization. This information is best obtained from a scatterplot of the scores following waking (or imagination) instructions and following attempted hypnotic induction. Throwing together for this purpose all subjects who have contributed one score under a waking or imagination condition and another following hypnotic induction, yields a total of 129 subjects from the two experiments. A scatterplot of the relationship between waking response and hypnotic response is given in Figure 2. The overall correlation is $r = .63$; the plot is not significantly curvilinear, the value of eta being .67, not significantly different from r. Even so, the scatterplot is somewhat triangular, with cases falling in three of the four quadrants as shown. What this means is that many who fall into the low-responsive group in tests of waking suggestions are indeed quite hypnotizable following a standard induction. These are the 34 cases in the upper-left quadrant. There are also a substantial number who are quite susceptible in the waking state, though many of them go further under hypnosis (the 30 cases in the upper right quadrant). It is of interest that 11 of these scored higher under hypnosis, 11 of these scored as high in waking as in hypnosis, and 7 scored slightly lower in the hypnotic test. The absent quadrant is the lower-righthand one, which means that those high in waking suggestion are never found low in hypnotic suggestibility. This is a statistical justification for selecting good hypnotic subjects by means of waking hypnosis tests; some good subjects will be missed, but no very poor ones will be chosen. Those whose suggestibility is unaffected by hypnosis lie in the diagonals; as expected, more of these are poorer subjects (20 scoring 0–4), but there are the 11 high-scoring ones also (scoring 5–10).

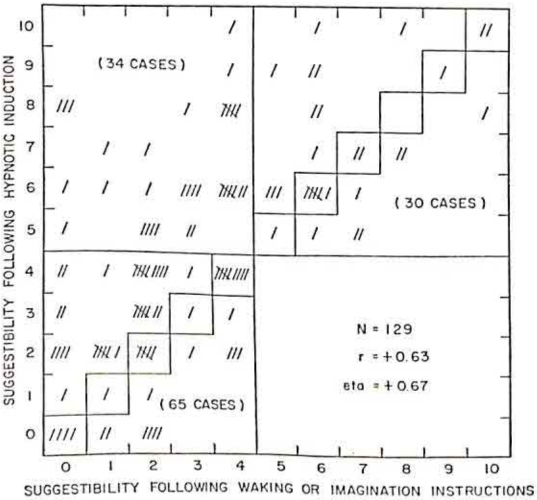


FIG. 2. Scatterplot of suggestibility scores following waking or imagination instructions. (The scores all fall within three of the four quadrants.)

Many of the contradictory statements about the effectiveness of hypnotic induction can be understood in the light of Figure 2. Depending upon the manner in which subjects are selected, and what kind of statistical treatment is employed, it is easy to demonstrate either that induction has a substantial effect or almost none at all. The overall finding is, of course, that the effect of hypnotic induction is very real for a number of subjects. Thus for the 99 subjects who score 0–4 in waking or imagination, 35 gained under hypnosis, many of them substantially, 20 remained the same, and 12 decreased a point or two.

DISCUSSION

The results support the usual belief that following hypnotic induction there is a statistically significant increase in responsiveness to suggestions of the kind familiar within hypnotic lore, although the amount of this increase, in agreement with Hull's early findings, is small. The reasons for the small average gains have been clarified by showing that many low-scoring subjects change very little as a consequence of induction, while others yield high scores on waking suggestion tests, and hence also gain little under hypnosis. Unless methods of investigation permit the use of the subject as his own control, these small gains are difficult to detect, and hence may lead to the conclusion of no difference, the conclusion frequently reached in recent publications by Barber and his associates. Treatment differences reached significance for randomly selected groups in only one of the comparisons in the present experiment (Table 3) in which each of the three groups contained 30 subjects, and even then one of the imagination groups was not significantly different from the hypnotic group. When subjects tested under these same imagination conditions were later tested following hypnotic injunction, they went on to a significant increase in their scores. The basic data are all therefore consistent with the findings of others, but the evidence refutes the interpretation of lack of influence of hypnotic induction.

The most serious criticism of experiments in which subjects serve as their own controls is that they may temper their performances in waking conditions in order to enhance them in hypnosis, as in the experiments of Scharf and Zamansky (1963) and Zamansky, Scharf, and Brightbill (1964). The current authors' experimental arrangements permit a number of comparisons to test the presence of this order-of-treatments effect. If it is present, it is not clearly demonstrable, for individual differences in susceptibility across groups are far more important. For example, the means in Table 5 can be studied for these effects. Consider Group 1, with imagination in the first session, hypnosis in the second, as compared with Group 3, with hypnosis first and imagination second. The hypothesis is that the imagination group would hold back on the first day for Group 1, probably not on the second day for Group 3. The means are in that order: 2.80 for Group 1 and 3.53 for Group 3. But the same argument would hold that both groups should enhance their hypnotic performances. It turns out that Group 3 has a mean for hypnosis of 5.03 and Group 1 a mean of 3.70, a little higher than the imagination mean for Group 3. An explanation more plausible than the holding-back hypothesis is that the subjects of Group 1 were merely less susceptible than those of Group 2; otherwise there is no reason for their low hypnosis scores on the second day. In Table 6 there is one comparison that might fit the expectation hypothesis, the results for the IE-IE-HYP group. On the assumption that these subjects, expecting hypnosis through imagination, and not expecting to be formally hypnotized at all, gave all they had in their 2 days, then the failure to gain when brought back for hypnosis on an unexpected third day is understandable according to the order-of-treatments hypothesis. The fact that their best is so relatively poor (mean = 3.55), weakens this interpretation. Probably the most telling argument against the order-of-treatments hypothesis is that from the correlation between state reports and responsiveness to suggestions. Those subjects who felt themselves to be hypnotized were the ones, *regardless of instructions*, or order of treatment, who yielded the higher suggestibility test scores. To argue that they *felt* hypnotized only because they expected to *become* hypnotized is merely to assert that there is in fact some responsiveness to suggestion involved in hypnosis; to deny responsiveness to suggestion would make the procedure of hypnosis meaningless. This constraint is not implied in looking for effects due to order of treatments.

A serious problem is raised in making responsiveness to suggestion a criterion of hypnosis, for it is evident that some subjects, not hypnotized, can yield high responses to suggestion. The state reports show that this is not too serious a matter because of the substantial correlations between state reports and responsiveness to suggestions, even when the state reports are made after induction, and before the suggestion tests are administered. The fact that the regression lines are distinguishable, even for correlations of about the same size (Figure 1), means that there is a difference, however, between the subject responding to suggestions following hypnotic induction and without such an induction. It would be desirable to develop some measure of the extent to which a person is actually hypnotized other than measures of his responsiveness to standard suggestions; it is quite possible that in the deeper stages of hypnosis responsiveness to some suggestions will disappear, without this meaning any weakening of the hypnosis. The ordinary tests of responsiveness to suggestions following standard hypnotic inductions are still quite satisfactory as measures of individual differences, and their validity as measures of individuality is supported by the correlations between waking and hypnotic responsiveness as well as by the correlations between responsiveness on these tests and state reports. This does not preclude further refinements, however, that would go beyond suggestibility, and might get at some of the other important aspects of the hypnotic state.

REFERENCES

BARBER, T. X., & CALVERLEY, D. S. "Hypnotic behavior" as a function of task motivation. *Journal of Psychology,* 1962, **54**, 363–389.

BARBER, T. X., & CALVERLEY, D. S. The relative effectiveness of task-motivating instructions and trance-induction procedures in the production of "hypnotic-like" behaviors. *Journal of Nervous and Mental Disease,* 1963, **137**, 107–116.

BARBER, T. X., & GLASS, L. B. Significant factors in hypnotic behavior. *Journal of Abnormal and Social Psychology,* 1962, **64**, 222–228.

HATFIELD, E. C. The validity of the LeCron method of evaluating hypnotic depth. *International Journal of Clinical and Experimental Hypnosis,* 1961, **9**, 215–221.

HILGARD, E. R., WEITZENHOFFER, A. M., LANDES, J., & MOORE, R. K. The distribution of susceptibility to hypnosis in a student population: A study using the Stanford Hypnotic Susceptibility Scale. *Psychological Monographs,* 1961, **75** (8, Whole No. 512).

HULL, C. L. *Hypnosis and suggestibility: An experimental approach.* New York: Appleton-Century, 1933.

LECRON, L. M. A method of measuring the depth of hypnosis. *Journal of Clinical and Experimental Hypnosis,* 1953, **1**, 4–7.

MCNEMAR, Q. *Psychological statistics.* (3rd ed.) New York: Wiley, 1962.

SCHARF, B., & ZAMANSKY, H. S. Reduction of word-recognition threshold under hypnosis. *Perceptual & Motor Skills,* 1963, **17**, 499–510.

SIEGEL, S. *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

TART, C. T. Hypnotic depth and basal skin resistance. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11**, 81–92.

TROFFER, S. H., & TART, C. T. Experimenter bias in hypnotist performance. *Science,* 1964, **145**, 1330–1331.

WEITZENHOFFER, A. M. *Hypnotism: An objective study in suggestibility.* New York: Wiley, 1953.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Form C.* Palo Alto, Calif.: Consulting Psychologists Press, 1962.

WEITZENHOFFER, A. M., & SJOBERG, B. M., JR. Suggestibility with and without "induction of hypnosis." *Journal of Nervous and Mental Disease,* 1961, **132**, 204–220.

ZAMANSKY, H. S., SCHARF, B., & BRIGHTBILL, R. The effect of expectancy for hypnosis on prehypnotic performance. *Journal of Personality,* 1964, **32**, 236–248.

(Received May 26, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 209–214

SOCIAL COMPETENCE AND PSYCHIATRIC DIAGNOSIS[1]

LESLIE PHILLIPS, INGE K. BROVERMAN
Research Institute of Life Sciences, Worcester State Hospital, Clark University
AND EDWARD ZIGLER
Yale University

It was hypothesized that symptoms reflect a patient's social competence or maturity level and that this maturity level is related to the diagnosis he receives. Symptom scales based on empirical relations between symptoms and competence levels were constructed for 3 spheres of functioning: thought, affect, and action. Results obtained with 504 psychiatric patients showed that (a) placement on each symptom scale was significantly related to diagnosis; and (b) patients who could be rated on all 3 scales showed consistent placement across scales. This latter finding was interpreted as indicating that the maturity dimension is a pervasive one, thus lending further support to the view that maturity plays an important role in the process of diagnosis.

A number of papers evaluating the reliability of our conventional system of psychiatric diagnosis (Foulds, 1955; Hunt, Wittson, & Hunt, 1953; Schmidt & Fonda, 1956; Seeman, 1953; Zigler & Phillips, 1961b) have indicated that as long as diagnosis is confined to broad diagnostic categories the system is modestly reliable. However, although clinicians show some consistency in placing patients into the major nosological categories, the process underlying this reliability remains something of a mystery.

When forced to be explicit about the process they use, most clinicians assert that they make their diagnosis on the basis of the symptomatic picture presented by the patient. This strikes one as being perfectly reasonable, since the classificatory principle for the diagnostic system currently in use is supposedly symptom manifestation. If symptoms do indeed define diagnosis, then any particular symptom or combination of symptoms should occur predominantly in a single diagnostic category. Such, however, is not the case. Several studies—see Freudenberg and Robertson (1956), Zigler and Phillips (1961c)—have indicated that although the occurrence of certain symptoms tends to be associated with particular diagnostic categories, many symptoms are related to more than one diagnosis. Furthermore, even in those cases where a statistically significant relationship exists between a symptom and a diagnostic class, the magnitude of this relationship is generally quite small. As Freudenberg and Robertson (1956) have noted, this minimal relationship between symptoms and diagnostic categories is not appreciably altered by employing complex combinations of symptoms rather than single ones. Thus the question is raised as to why psychiatric classification is as reliable as it is when symptoms are so minimally related to diagnosis. Thorne (1953) proposed a solution to this problem in suggesting that the experienced clinician does not diagnose on the basis of textbook symptoms but primarily on the basis of "the feel of the case." Thus, the experienced diagnostician is seen as an analogue of the computer, a view reminiscent of that suggested by Eysenck (1953).

Drawing such a parallel reveals little about the nature of the clinician's "program," that is, the configuration which transcends symptomatology or those criteria which lie beyond the domain of psychiatry proper to which he is responding. One possibility is that while the clinician orients himself toward the symptoms presented by the patient, the actual process of diagnosis is mediated by the clinician's ordering these symptoms along certain dimensions. Some clue concerning the nature of one such dimension may be gleaned from that clinical folklore which characterizes the manic-depressive and the psychoneurotic as "mature," "of higher socio-economic class," "phallic," etc., in contrast to schizophrenics and to character disorders who are typically characterized as "immature," "regressed," "of lower socio-economic class," "oral," etc. This theoretical ordering of diagnostic categories along a dimension of maturity receives empirical support from the findings that diagnosis is, in fact, related to the general maturity level of the individual as measured by social competence indices which both include and go beyond those typically employed to define social class (Zigler & Phillips, 1961a).

---

[1] This investigation was supported by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (MH06369) from the National Institute of Mental Health, United States Public Health Service. This work was done in part at the Computation Center, Clark University, and the Computation Center of Massachusetts Institute of Technology, Cambridge.

Evidence that the individual's social competence or maturity level is a pervasive factor in a variety of psychopathological phenomena (e.g., incidence of mental disorder, gross symptomatic picture presented and prognosis) has now been presented in a series of papers (Phillips & Rabinovitch, 1958; Phillips & Zigler, 1961, 1964; Zigler & Phillips, 1960, 1961a, 1961c, 1961d, 1962). The centrality of this dimension has led the authors to propose that clinicians implicitly order symptoms along a dimension of maturity when making a diagnosis.

In the typical diagnostic procedure the clinician assesses the presence or absence of symptoms in various spheres of psychological functioning. The diagnostician closely examines the character of the patient's thought processes, the nature of his affect, as well as those acting-out behaviors which might have precipitated referral or hospitalization. Within each of these spheres of Thought, Affect, and Action the clinician attempts to discern behaviors which are considered symptomatic of particular mental disorders. A large proportion of the symptoms represented in any hospital population can, on an a priori basis, be divided into these three spheres. In the present study the authors will attempt to demonstrate that certain symptoms bear a systematic relation to maturity level as measured by achieved level of social competence. The inference is made that the diagnostician *takes into account* this maturity-symptom relation in making a diagnosis.

The plan of investigation was based on the following considerations. Independent of diagnostic considerations, any symptom can be examined for its empirical relationship to maturity as measured by premorbid social competence indices. Once this is done a symptom scale which employs only those symptoms which represent discrete points along the social competence dimension can be devised for each sphere of psychological functioning. The specific prediction would then be tested that the diagnosis which a patient receives is a function of where his symptoms are located on the symptom scales constructed for each of the three spheres.

Furthermore, if the maturity dimension does pervade all areas of functioning then one would expect a positive relation among the three symptom scales. That is, patients who manifest symptoms indicative of relatively high maturity in one sphere should tend to manifest mature symptoms in the other two spheres as well. The evidence for such a general maturity factor would give further credence to the view that the clinician should and probably does respond to such a factor when making a diagnosis.

PROCEDURE

Subjects

This study was based on the examination of the case histories of 504 patients, 287 males and 217 females, admitted to Worcester State Hospital during a 12-year period (1945–1957). The particular case histories employed were those of patients referred to the Hospital Psychology Department for psychological testing, and who were eventually diagnosed as suffering from a functional disorder. While this sample includes a wide variety of cases, extremely deteriorated or very agitated patients are seldom referred for psychological evaluation and are not adequately represented in this sample.

Diagnosis

The diagnosis ascribed to each patient in the study was that psychiatric classification agreed upon at a diagnostic staff conference and noted on the patient's hospital record as his primary diagnosis. The diagnostic classification of mental illness used by Worcester State Hospital during this period was based on the following sources: *Standard Classified Nomenclature of Disease* (1945–1952); and, *Diagnostic and Statistical Manual of Mental Disorder* (1953–1957).

The patients in the present study were categorized into four major diagnostic groups: manic-depressive, schizophrenic, psychoneurotic, and character disorder.

The composition of each of these groups was as follows:

1. Manic-depressive here was defined to include the diagnoses of involutional psychotic reaction; manic-depressive, manic; manic-depressive, depressive; manic-depressive, other; and psychotic depressive reaction ($N = 47$).

2. The schizophrenic population included those with diagnoses of the following types of schizophrenic reactions: simple, hebephrenic, catatonic, paranoid, undifferentiated, and schizo-affective ($N = 182$).

3. Psychoneurotic included diagnoses of the following types of psychoneurotic reactions: anxiety, dissociative, conversion, phobic, obsessive-compulsive, depressive, and other ($N = 98$).

4. Character disorder included the diagnoses of personality-pattern disturbance—inadequate, schizoid, cyclothymic, or paranoid type; personality-trait disturbance—emotionally unstable, passive-aggressive, compulsive personality, or other type; and, sociopathic disturbances of the following types: antisocial; dyssocial; sexual deviation; addiction, alcoholism; and, addiction, drug addiction ($N = 177$).

The decision to employ such broad classifications is dictated by the fact that the narrower the diagnostic category the less reliable it is (Zigler & Phillips, 1961).

Social Competence

In earlier studies the variables of education, occupation, intelligence, employment history, marital status, and age were used as social competence indices. Each of these variables was divided into a number of categories which were conceptualized as representing steps along a maturity continuum (Zigler & Phillips, 1960). The number of categories in each variable ranged from a low of three for marital status, to a high of six for education and intelligence. In the present study, however, only the variables containing five or more categories, that is, occupation, intelligence, and education, were used to compute a patient's maturity level, since these items were thought to have sufficiently fine gradations to permit a discrete ordering of symptoms.

The three variables and their categories from low to high are as follows:

1. Occupation—*The Dictionary of Occupational Titles* (1949) was employed to place each patient's occupation into the categories of unskilled; semiskilled or service; clerical or sales; skilled; and professional or managerial.

2. Intelligence—IQs obtained on standard intelligence tests of 69 or less; 70–79; 80–89; 90–119; 120–129; 130 or more.

3. Education—none or some grades; finished grade school; some high school; finished high school; some college; finished college or more.

Each patient received a score on each of the three social competence indices based on case-history information. In order to make the three indices additive, a score between 0 and 100 was assigned for each index giving the three indices approximately the same means and ranges. For the six-category variables, education and intelligence, the scores assigned from low to high were 8, 25, 41, 58, 75, 92. For the five-category variable, occupation, the scores from low to high were 10, 30, 50, 70, and 90. The overall maturity score for each patient was the mean of the scores obtained on the three variables. This averaging procedure was required since data were not available for every patient on all three variables.

In order to check on the extent to which this maturity score represented the previously used index of social maturity a Pearson product-moment correlation was performed between the two measures. The resulting correlation of .68 was sufficiently high to consider the maturity score based on three variables as strongly related to the previously used social competence measure based on six variables. However, given a correlation of this moderate magnitude, the two indices cannot be considered identical.

Symptoms

A symptom in the present study, as in earlier studies (Phillips & Rabinovitch, 1958; Zigler & Phillips, 1960), refers to the description of a patient's behavior by a psychiatrist at the time of initial institutional contact, or the description of behavior presented by referring physicians as the primary reason for hospitalization. The symptoms include specific actions, for example, assault; general behavior patterns, for example, irresponsible behavior; thoughts, for example, sexual preoccupations; somatic reactions, for example, headaches; and general affect states, for example, being tense. Examination of the 504 case records resulted in the tabulation of the following 46 discrete-presenting symptoms:

All addictions except drinking
Amnesia
Apathetic
Assaultive
Blackouts
Bizarre ideas or delusions
Bodily complaint
Charged with murder
Compulsions
Convulsions
Depersonalization
Doesn't eat
Depressed
Drinking
Eats a lot
Euphoria
Feels going crazy
Fear of own hostile impulses
Feels perverted
Feelings of sexual inadequacy
Fire setting
Headaches
Hallucinations
Homosexuality
Insomnia
Irresponsible behavior
Lying
Maniacal outbursts
Memory impairment
Mood swings
Obsessions
Perplexed
Phobias
Psychosomatic disorders
Rape
Robbery
Sexual deviations, except homosexuality
Sexual preoccupation
Self-depreciatory
Suicidal attempt
Suicidal ideas
Suspicious
Temper outbursts
Tense
Threatens assault
Withdrawn

On rational grounds the following symptoms were classified as Thought symptoms:

- Bizarre ideas or delusions
- Depersonalization
- Feels going crazy
- Fear of own hostile impulses
- Feels perverted
- Feelings of sexual inadequacy
- Hallucinations
- Obsessions
- Perplexed
- Phobias
- Self-depreciatory
- Sexual preoccupations
- Suicidal ideas
- Suspicious
- Threatens assault

As Affect symptoms:

- Apathetic
- Depressed
- Euphoria
- Maniacal outbursts
- Mood swings
- Temper outbursts
- Tense

As Action symptoms:

- All addictions except drinking
- Assaultive
- Charged with murder
- Compulsions
- Doesn't eat
- Drinking
- Eats a lot
- Fire setting
- Homosexuality
- Irresponsible behavior
- Rape
- Robbery
- Sexual deviations, except homosexuality
- Suicidal attempt

Symptoms often do not lend themselves to unambiguous placement in any one category. The classification presented here represents an effort to assign symptoms reasonably into one of three categories. The authors were unable to place 10 symptoms and so removed them from further consideration.

Construction of Thought, Affect, and Action Scales

An effort was made to obtain social competence values for each symptom and combination of two symptoms within a symptom category, free from the presence of other symptoms within that category. Therefore, only those patients were selected who had at least one, but no more than two symptoms within any one of the three symptom categories. One hundred fifty-four, 264, and 187 met this criterion for the Thought, Affect, and Action categories, respectively. (It should be noted that a single case could fall into more than one category.) Of the total of 504 patients in the initial sample, 383 (about 76%) met the criteria for inclusion on at least one symptom scale.

In the construction of the Thought symptom scale only those cases were employed that had no more than two thought symptoms. In order to obtain a social competence score for each thought symptom and every possible combination of two symptoms, the overall social competence scores of the patients exhibiting that symptom or combination of symptoms were averaged. A series of t tests was then run between the social competence scores associated with each symptom and symptom combination. For the final Thought scale those symptoms and symptom combinations were selected that: (*a*) represented significantly different points along the social competence continuum; and (*b*) occurred in a sufficient number of cases to make them suitable for the type of statistical treatment to be employed. The final Thought scale was composed of symptoms representing three distinct levels of competence. The symptoms and symptom combinations as well as their respective mean levels of competence from high to low are presented in Table 1. This same procedure was employed in the construction of the Affect scale and the Action scale which are also presented in Table 1.

RESULTS AND DISCUSSION

Table 2 presents the number of patients in each diagnostic category manifesting symptoms or symptom combinations at each competence level for each symptom scale. (It should be noted that a single patient may be included in more than one scale.) As predicted, each scale was found to be significantly related to diagnosis.

TABLE 1

SOCIAL COMPETENCE LEVEL AND SYMPTOM SCALES

Scale level	Symptoms	Mean competence score
Thought scale		
1.	suicidal ideas; self-depreciatory	50.75
2.	suspicious; suspicious and threatens assault; perplexed	44.17
3.	hallucinations; hallucinations and suspicious[a]	41.56
Affect scale		
1.	depressed; depressed and tense	48.30
2.	tense; apathetic; apathetic and depressed	42.98
3.	temper outbursts; temper outbursts and tense	33.84
Action scale[b]		
1.	suicidal attempt; suicidal attempt and drinking	50.62
2.	robbery; drinking; sexual deviation, except homosexuality; assaultive; assaultive and drinking	43.71

---

[a] It will be noted that "suspicious" appears at more than one level. If it is found as the only thought symptom or appears in combination with "threatens assault" it is at the second level of competence. If it occurs in combination with hallucinations it is at the third level.

[b] Action scale—Only two levels of competence could be distinguished for symptoms used in the construction of the Action scale.

TABLE 2

RELATIONSHIPS BETWEEN SYMPTOM SCALES AND DIAGNOSIS

Symptom scale	Diagnosis			
	Manic-depressive	Schizo-phrenic	Psycho-neurotic	Character disorder
Thought				
Level 1	5 (4.4)	5 (20.1)	19 (9.9)	19 (13.4)
Level 2	7 (5.6)	31 (25.6)	9 (12.6)	14 (17.0)
Level 3	2 (3.8)	27 (17.2)	3 (8.4)	9 (11.4)
	$\chi^2 = 35.58$ $p < .001$			
Affect				
Level 1	21 (14.0)	39 (51.1)	45 (35.9)	46 (49.9)
Level 2	0 (6.4)	37 (23.3)	12 (16.4)	20 (22.8)
Level 3	2 (2.5)	8 (9.4)	2 (6.6)	16 (9.2)
	$\chi^2 = 33.49$ $p < .001$			
Action				
Level 1	11 (8.4)	16 (17.3)	24 (15.1)	22 (32.0)
Level 2	9 (11.5)	25 (23.6)	12 (20.8)	54 (43.9)
	$\chi^2 = 15.94$ $p < .005$			

Note.—Frequencies expected by chance alone presented in parenthesis.

In respect to the Thought scale, psychoneurotics, and to a lesser extent, character disorders, were characterized by relatively high-level thought symptoms while the reverse was true for schizophrenics. Findings with the Affect scale indicated that manic-depressives and psychoneurotics evidenced high-level affect symptoms while schizophrenics and character disorders were characterized by lower-level affect symptoms. On the Action scale, psychoneurotics manifested high-level symptoms while character disorders had low-level symptoms. In general then, psychoneurotics and manic-depressives were found to have symptoms indicative of higher levels of social maturity than those found in schizophrenics and character disorders. However, differences within these pairs of diagnosis should be noted. The manic-depressive's higher level of social maturity is evidenced most clearly in affect while the psychoneurotic tends to manifest high-level symptoms in all three areas. Schizophrenics express their immaturity primarily in the sphere of thought and character disorders in the sphere of affect.

Table 3 involves only those patients who appear on at least two symptom scales and shows the relationships obtained between the levels of the three scales. As predicted, positive relations were found between all three symptom scales.

In general, the findings lend support to the authors' premise that a patient's level of social maturity finds expression in his symptomatology and that this combination of maturity and symptom expression is an important factor in the diagnosis that he receives. That the maturity dimension is a pervasive one is attested to by the findings presented in Table 3, which indicate that the symptoms of a patient across spheres of functioning are not idiosyncratic, but rather that they are a consistent expression of the patient's maturity level. (The reader should remember that only a minority of patients are represented in Table 3, and that the symptom scale levels were computed primarily on the basis of social competence scores of a much larger pool of patients than those represented in Table 3. Had the scales been constructed only on those patients who are represented on more than one scale, then a positive relation between scales would not only be inevitable but also close to perfect.)

A remaining question of interest is why the rather simple symptom scales in the present study were found to be relatively highly related to diagnosis while discrete symptoms, as well as other combinations of symptoms, were found to be minimally related to diagnostic categories in earlier studies (Freudenberg & Robertson, 1956; Wittenborn, Holzberg, & Simon, 1953; Zigler & Phillips,

TABLE 3

RELATIONSHIPS BETWEEN THE THREE SYMPTOM SCALES

Symptom scale	Symptom scale			
	Affect		Action	
	Level 1	Level 2 & 3	Level 1	Level 2
Thought				
Level 1	26 (20.8)	8 (13.1)	17 (10.0)	3 (10.0)
Level 2	11 (13.5)	11 (8.5)	10 (13.0)	16 (13.0)
Level 3	9 (11.6)	10 (7.3)	4 (8.0)	12 (8.0)
	$\chi^2 = 6.04$ $p < .05$		$\chi^2 = 15.18$ $p < .001$	
Action				
Level 1	30 (21.3)	6 (14.7)		
Level 2	18 (26.7)	27 (18.3)		
	$\chi^2 = 15.55$ $p < .001$			

1961c). In the earlier studies either discrete or certain selected combinations of symptoms were related to diagnosis, while in the present study a theoretical framework and empirical evidence were employed to construct symptom scales consisting of a number of symptoms and symptom combinations. Thus, in the present study an atomistic, atheoretical approach to relations between symptoms and diagnosis was eschewed in favor of a theoretical one in which a number of symptoms and symptom combinations representative of the same maturity level were conceptualized as functionally equivalent indicators of a diagnosis. Furthermore, only those symptoms were employed which represented discrete points along a social competence continuum and which were pathognomonic in nature, that is, actually led to hospitalization. It may be that the large number of symptoms excluded are those that create "noise" in the diagnostician's system and which consequently, may have led to the generally low relations found between symptomatology and diagnosis in previous studies.

It is of considerable interest that even with the exclusion of so many of the original symptoms and the rather stringent criteria used for including a patient in the final study, approximately 76% of the patients included in the present study were classifiable on at least one symptom scale. This indicates that the study dealt neither with a very esoteric group of symptoms nor a very select sample of patients.

One final note is in order. Though the relations found between the symptom scales and the diagnostic categories were highly significant, they were far from perfect. This is what one would expect if, as has been hypothesized, diagnosticians employ something akin to these scales in an implicit way when making a diagnosis. To the extent that such a procedure is employed by diagnosticians, the diagnostic system will be a somewhat reliable one. It would follow, however, that the reliability of psychiatric diagnosis might be improved considerably by the explicit incorporation of the maturity dimension into the classification process.

REFERENCES

EYSENCK, H. The logical basis of factor analysis. *American Psychologist,* 1953, **8**, 105–113.

FOULDS, G. The reliability of psychiatric, and the validity of psychological diagnosis. *Journal of Mental Science,* 1955, **101**, 851–862.

FREUDENBERG, R., & ROBERTSON, J. Symptoms in relation to psychiatric diagnosis and treatment. *Archives of Neurology and Psychiatry,* 1956, **76**, 14–22.

HUNT, W., WITTSON, C., & HUNT, E. A theoretical and practical analysis of the diagnostic process. In P. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis.* New York: Grune & Stratton, 1953. Pp. 53–65.

PHILLIPS, L., & RABINOVITCH, M. Social role and patterns of symptomatic behaviors. *Journal of Abnormal and Social Psychology,* 1958, **57**, 181–186.

PHILLIPS, L., & ZIGLER, E. Social competence: The action-thought parameter and vicariousness in normal and pathological behaviors. *Journal of Abnormal and Social Psychology,* 1961, **63**, 137–146.

PHILLIPS, L., & ZIGLER, E. Role orientation, the action-thought dimension, and outcome in psychiatric disorder. *Journal of Abnormal and Social Psychology,* 1964, **68**, 381–389.

SCHMIDT, H., & FONDA, C. The reliability of psychiatric diagnosis: A new look. *Journal of Abnormal and Social Psychology,* 1956, **52**, 262–267.

SEEMAN, W. Psychiatric diagnosis: An investigation of interperson-reliability after didactic instruction. *Journal of Nervous and Mental Disease,* 1953, **118**, 541–544.

THORNE, F. Back to fundamentals. *Journal of Clinical Psychology,* 1953, **9**, 89–91.

WITTENBORN, J., HOLZBERG, J., & SIMON, B. Symptom correlates for descriptive diagnosis. *Genetic Psychology Monographs,* 1953, **47**, 237–301.

ZIGLER, E., & PHILLIPS, L. Social effectiveness and symptomatic behaviors. *Journal of Abnormal and Social Psychology,* 1960, **61**, 231–238.

ZIGLER, E., & PHILLIPS, L. Case history data and psychiatric diagnosis. *Journal of Consulting Psychology,* 1961, **25**, 458. (a)

ZIGLER, E., & PHILLIPS, L. Psychiatric diagnosis: A critique. *Journal of Abnormal and Social Psychology,* 1961, **63**, 607–618. (b)

ZIGLER, E., & PHILLIPS, L. Psychiatric diagnosis and symptomatology. *Journal of Abnormal and Social Psychology,* 1961, **63**, 69–75. (c)

ZIGLER, E., & PHILLIPS, L. Social competence and outcome in psychiatric disorder. *Journal of Abnormal and Social Psychology,* 1961, **63**, 264–271. (d)

ZIGLER, E., & PHILLIPS, L. Social competence and the process-reactive distinction in psychopathology. *Journal of Abnormal and Social Psychology,* 1962, **65**, 215–222.

(Received May 26, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 215–223

STRESS SITUATIONS, BIRTH ORDER, AND PSYCHOLOGICAL SYMPTOMS[1]

BARBARA SNELL DOHRENWEND AND BRUCE P. DOHRENWEND
The City College of the City University of New York *Columbia University*[2]

Recent conflicting findings raise the issue of whether observed differences in levels of psychological distress of 1st and later borns are a function of their contemporary social situation, or of basic personality defect in 1 or the other group. Hypotheses which might explain conflicting findings are derived from Schachter's work: among Ss exposed to stressors, 1st borns will show more symptoms when social isolation is enforced, and later borns when social interaction is enforced. These hypotheses are tested in a socially isolating sensory deprivation study, and in a community sample including crowded slum dwellers. Both hypotheses are supported.

Since Schachter (1959, pp. 46, 100) discovered relations between subjects' birth order and their affiliative responses in anxiety provoking situations, a latent issue has developed concerning the nature of the association between birth order and symptoms of psychological distress. While Schachter showed that differences in levels of anxiety between first- and later-born subjects were specific to anxiety provoking situations, other researchers have reported associations between birth order and measures assumed to indicate persistent psychological disorder (e.g., Altus, 1959; Greenberg, Guerino, Lashen, Meyer, & Piskowski, 1963; Prakash & Srivastava, 1963; Schooler, 1961; Vogel & Lauterback, 1963; Vuyk, 1963). The issue that emerges is whether observed differences in levels of psychological distress of first- and later-born persons are a function of their contemporary social situation, or of basic personality defect in one or the other group.

It is possible of course that each of these propositions is correct within limits. If this position were to be accepted, however, one would still be faced with the problems of explaining the inconsistencies between studies which show, for example, higher rates of persistent disorder among later born or last born (e.g., Schooler, 1961; Vuyk, 1963) as against higher rates among first born,[3] and less disorder among last born (e.g., Greenberg et al., 1963). Considerations of parsimony, therefore, favor Schachter's hypothesis, since it is not only consistent within itself, but might provide situational explanations of the varied findings concerning relations between birth order and rate of psychological disorder.

Schachter (1959) and others (Sarnoff & Zimbardo, 1961) demonstrated that under anxiety arousing conditions first-born subjects (including only children) tended to seek the company of others. Furthermore, Wrightsman (1960) showed that in first-born subjects anxiety was reduced more when they were in the company of others than when they were alone.

The results for later-born subjects, while less consistent, suggest the complementary reaction. In experimental studies, later-born subjects tended to be indifferent to the company of others when anxious (e.g., Schachter, 1959). Although Wrightsman's (1960) later-born subjects showed no more anxiety reduction when alone than when with others, Schachter (1959, pp. 69–70) presented findings that, among veterans applying for psychotherapy—presumably an anxious group—later borns were more likely than first borns to withdraw before the fourth therapeutic session, suggesting that they did not find this intense social interaction anxiety reducing. The inconsistency between this finding concerning later borns and Wrightsman's may be explained by the brief and permissive nature of Wrightsman's social interaction. His subjects were exposed for only 5 minutes to the company of three strangers with whom they might, but were not required to, talk. Thus, on the basis of the evidence available, it seems that later-born subjects may reduce anxiety more effectively when they are not forced to interact with others than when they are.

---

[1] This paper is an expanded version of a report presented at the 1965 convention of the American Psychological Association in Chicago. The work was supported in part by Grants MH 07327, MH 07328, and MH 10328 from the National Institute of Mental Health, United States Public Health Service. We would like to thank Gertrude R. Schmeidler, Sol Feldstein, and Joyce Plosky for their help in securing the data on effects of sensory deprivation, and John Colombotos and Gladys Egri for their advice in classifying stressor events.

[2] Community Psychiatry Research Unit, 128 Ft. Washington Avenue, New York, New York 10032.

[3] Leonard F. Solomon and Ronald L. Nuttall. Personal communication. February 1965.

These findings can be analyzed to explain varied relations between birth order and symptoms of psychological distress. The analysis is based on two assumptions, both consistent with clinical experience as well as the research literature: (*a*) When anxiety is aroused in an individual he is motivated to reduce it; (*b*) Prolonged anxiety arousal produces symptoms of psychological disturbance (cf. Wilson, 1963). These assumptions, taken together with Schachter's results concerning the different social situations in which first and later borns are best able to reduce anxiety, lead to predictions about the circumstances in which symptoms of psychological distress should be higher among first borns, higher among later borns, and equal for the two groups.

All cases begin with the presence of anxiety-arousing conditions. It is then predicted that:

1. Insofar as social isolation is forced on individuals exposed to anxiety-arousing conditions, first borns will show more symptoms of psychological distress.
2. Insofar as social interaction is forced on individuals exposed to anxiety-arousing conditions, later borns will show more symptoms of psychological distress.
3. Insofar as individuals exposed to anxiety-arousing conditions are free to choose social isolation or social interaction, both first- and later-born persons will seek the condition which permits them to reduce anxiety, and neither group will show a higher level of symptoms of psychological disturbance.

This paper draws on two studies to investigate the above predictions. A sensory deprivation study is used to test the first prediction concerning the effects of enforced social isolation. A panel survey of a community sample, including crowded slum dwellers, is used to test the prediction concerning the effects of forced social interaction, together with a corollary of the third prediction concerning the situation of free choice. The measure of the dependent variable, psychological symptoms, is the same in both studies.

MEASUREMENT OF PSYCHOLOGICAL SYMPTOMS

The measure of symptoms is a 22-item screening instrument from the Midtown Study. Midtown researchers reported that each of these 22 items discriminated at the .01 level or better between a group of patients and persons rated "well" on the basis of a half-hour psychiatric interview. In addition, a score of four or more symptoms was found to discriminate between persons rated impaired as against those rated unimpaired by psychiatrists making judgments based on interview protocols (Langner, 1962). These symptoms are, therefore, indicators of psychological disturbance as defined by experienced clinicians.

SENSORY DEPRIVATION STUDY

Procedure

Sixteen college student volunteers, 11 males and 5 females, were paid $50 each for an 8-hour sensory deprivation session plus preliminary testing. During sensory deprivation the subject lay on a cot in a nearly sound-proof, lighted chamber with padded gloves and sleeves over his hands and arms, half Ping-Pong balls over his eyes, and his ears covered with earphones through which white noise was transmitted. One meal was provided, and toilet facilities were available in the deprivation room, so that the subject did not leave during the 8-hour period. Six to eight interruptions for tests involving a sentence or two of speech by the experimenter were the only breaks in the 8-hour sensory deprivation. The subject was encouraged to express his thoughts and feelings during deprivation. If he did not speak for a half-hour stretch, the microphone connected with his earphones was tapped as a prearranged signal that he should speak. This device also prevented extended sleep. The subjects' verbalizations were recorded on tape.

The announcement posted for recruiting subjects was designed to arouse anxiety about the experience, mentioning, for example, that "Some people have had very unpleasant experiences in the sensory deprivation situation." After a volunteer had been accepted he was told that he was to be paid $50, reinforcing his impression of the possible difficulties facing him.

The day before the deprivation session the subjects were tested and interviewed about a number of topics, including birth order and psychological symptoms. They were interviewed about the deprivation experience immediately after its completion, before leaving the deprivation chamber. A year to a year and a half after the sensory deprivation experience, all subjects who could be found were again contacted and interviewed about their usual psychological symptoms.

In the interview after the deprivation session the subjects were asked 20 of the 22 symptom items with reference to whether they experienced them during deprivation. Two of the 22 items were dropped in the postdeprivation interview because of difficulty in phrasing them so they could be referred to the deprivation period: "alone even among friends," and "the worrying type." Consequently, all symptom scores for the sensory deprivation subjects are based on the remaining 20 items rather than the full 22.

Results

In the predeprivation interview, the subjects were asked about the symptoms usually experienced. These symptom counts are shown in Table 1. Although there is a trend in the direction of higher symptoms among later-born subjects, the difference between the groups is not significant.

In the social isolation of the sensory deprivation situation, first-born subjects should have had difficulty reducing the anxiety which had been aroused. Hence they should show elevated symptom levels compared to their usual level. Later-born subjects, on the other hand, should have found that total removal from social interaction was conducive to anxiety reduction, and hence should show reduced symptom levels. The distribution of changes in Table 2, for which the exact probability is less than .05, supports this prediction.

TABLE 1

NUMBER OF PSYCHOLOGICAL SYMPTOMS REPORTED AS USUAL PRIOR TO SENSORY DEPRIVATION

Symptoms	Number of subjects	
	First born	Later born
0	2	1
1	2	2
2	3	2
3	0	3
4	0	1
	$p^a > .05$	

[a] Exact probability when symptom levels are divided into two categories: 0–2, 3 or more; this cutting point was chosen on inspection as the one which would produce the minimum probability.

TABLE 2

NUMBER OF PSYCHOLOGICAL SYMPTOMS EXPERIENCED BY FIRST- AND LATER-BORN SUBJECTS DURING SENSORY DEPRIVATION COMPARED TO USUAL LEVELS OF SYMPTOMS BEFORE AND AFTER

	Number of subjects	
	First born	Later born
Relation of level of symptoms during sensory deprivation to level reported as usual before deprivation		
Higher	4	2
Same	3	2
Lower	0	5
	$p = .045^a$	
Relation of level of symptoms reported as usual approximately 1 year after sensory deprivation to level during deprivation		
Higher	0	1
Same	0	2
Lower	5	2
	$p = .38^a$	

[a] Excluding "Same."

Comparison of symptoms during sensory deprivation with the number reported as usual about a year later strengthens the inference that the symptom levels during sensory deprivation were a function of the enforced isolation. Since the subjects would rarely in their daily lives experience social isolation as extreme as that in the sensory deprivation study, we would expect reversal of the changes which occurred during deprivation. That is, the level of symptoms should decrease among first borns and increase among later borns. While the changes among later-born subjects are not consistent, those for first borns clearly are as predicted. Hence the evidence from this study indicates that anxiety arousal coupled with social isolation is sufficient to produce situationally specific symptoms of psychological distress in first borns.

PANEL SURVEY

Procedure

The research setting of the panel survey was Washington Heights, a section of New York City surrounding the Columbia Presbyterian Medical Center. In 1920, Washington Heights was evenly populated by three major ethnoreligious groups:

Anglo-Saxon, Dutch, or German Protestants; Irish Catholics; and East-European Jews. Throughout the 30s and 40s, there continued to be a steady influx of Irish and Jews, with corresponding decreases in the original Protestant population. During this period up until World War II, the area was predominantly middle class in character and was inhabited almost wholly by whites of European ancestry. With the war came great changes that are continuing into the present.

As is true of the entire city, there has been much movement out of the area on the part of higher-income families resettling in the suburbs. There has also been heavy in-migration by Negroes from the South and, since 1950, by Puerto Ricans. Today, these two ethnic groups make up the large majority of the families living in the southeastern quarter of the area, in the older, more deteriorated housing. This section, now inhabited for the most part by low-income families, is rapidly taking on the character of a slum.

In July 1961, interviews were completed with 1,710 respondents constituting a representative sample of people 21 years of age and older in Washington Heights (Elinson & Loewenstein, 1963). Of these, 1,023 were Irish, Jewish, Negro, or Puerto Rican adults between the ages of 21 and 59. The interview covered both demographic factors and certain aspects of psychiatric symptomatology, including the 22 Midtown Study items used in the sensory deprivation study.

In 1962, a probability subsample of 94 married couples and 26 single household heads (214 individuals in all) was chosen for reinterviews from the previous survey. The subsample was limited to respondents between the ages of 21 and 59 at the time of the first interview living in households headed by Irish, Jewish, Negro, or Puerto Rican males. Interviews were completed with the designated respondents in 69% of the households. An additional 12% had to be removed from the sample because we could not locate them, verified moves out of the state, death, etc. Respondents in 19% of the households, disproportionately Irish, refused to be interviewed after repeated call-backs. In all, interviews were obtained from 151 of the designated 214 respondents.

Both the first and second interviews included the 22 Midtown items. The second interview also included questions about changes between the two interviews, especially changes in work and physical health.

The relevant questions about intervening events were as follows:

> Let's see, a Columbia University interviewer first visited your family to ask about health matters in __________. That was about ______ months ago.
>
> Have you had any serious illness or health problems since then?
>
> (If yes:) What is (was) the matter?
>
> What were you doing most the past ______ months?
>
> (For males:) Working, looking for work, or something else?
>
> (For females:) Working, looking for work, keeping house, or something else?
>
> Has there been any change in your work or job status since the time of the first interview ______ months ago?
>
> (If yes:) What change was that? Why did you change?

The responses to these questions were classified into the following categories of intervening events.

Unfavorable events:

1. Poor physical health of respondent; that is, physical illness, injury, accident, etc., which seems current in negative consequences either because the respondent says so or because it seems inherently so.
2. Poor physical health of respondent's spouse.
3. (For male main earner:) Socioeconomic deterioration; that is, demotion, business failure, unemployment, etc.
4. (For female spouses:) Economic deterioration; that is, demotion, business failure, unemployment, etc., of male main earner.

Favorable events:

1. (For male main earners:) Socioeconomic improvement; that is, promotion, raise, "better job," etc.
2. (For female spouses:) Socioeconomic improvement; that is, promotion, raise, "better job," etc., for male main earner.
3. Birth of one or more children.

Unknown:

1. Events related to the above whose current consequences could not be ascertained.
2. Other events which were mentioned adventitiously, that is, not in response to specific questions.
3. No evidence of any intervening event.

The classification of these events was made independently by a psychiatrist and a sociologist who had no information about respondents' psychological symptoms. They disagreed on 14 of the 74 events classified by either one or the other in the above favorable and unfavorable categories. None of these disagreements was over whether an event was favorable or unfavorable.

Results

The panel survey will be used to test the hypothesis that later borns show more symptoms of psychological disturbance than first borns when forced, in anxiety-arousing conditions, to interact with others. In order to test this hypothesis two assumptions are made. The first is that the residents of crowded slums have less privacy from family

TABLE 3

NUMBER OF PSYCHOLOGICAL SYMPTOMS REPORTED IN TWO SUCCESSIVE INTERVIEWS BY RESPONDENTS CLASSIFIED ACCORDING TO DISADVANTAGED OR ADVANTAGED ETHNIC STATUS AND BIRTH ORDER

Symptoms	Ethnic status			
	Disadvantaged (Negro and Puerto Rican)		Advantaged (Irish and Jewish)	
	First born %	Later born %	First born %	Later born %
First interview				
Less than 4	94	65	85	76
4 or more	6	35	15	24
Total	100	100	100	100
Number of subjects	17	49	27	41
	$\chi^2 = 3.93$ $p < .05$		$\chi^2 = 0.42$ $p > .05$	
Second interview				
Less than 4	100	65	75	81
4 or more	0	35	25	19
Total	100	100	100	100
Number of subjects[a]	17	51	32	43
	$\chi^2 = 6.45$ $p < .05$		$\chi^2 = 0.15$ $p > .05$	

[a] The number of subjects is different on the second interview because 11 women were interviewed the second but not the first time, while one man and one woman who answered the symptom items on the first interview broke off before answering them on the second interview.

and neighbors than do residents of less-crowded middle-class residential areas. Secondly, it is assumed that anxiety-arousing conditions are common in daily urban living.

On the basis of these assumptions, it is predicted that later-born slum dwellers should find anxiety reduction particularly difficult and should, therefore, have more symptoms of psychological distress than first-born slum dwellers. Outside of the slums, where privacy is more available, later borns should show no more symptoms than first borns. Since the slum sections of the area where the survey was conducted are inhabited almost exclusively by Negroes and Puerto Ricans, it is in these two otherwise disparate groups that we should find relatively high levels of psychological symptoms among later borns compared to first borns. The non-Puerto Rican white residents of nearby middle-class areas should not show this contrast between first and later born.

Table 3 shows the proportions of first- and later-born subjects in disadvantaged and advantaged ethnic groups having less than four as against four or more symptoms. We see that, as predicted, later-born members of the disadvantaged groups have more symptoms than first-born members of these groups. The more advantaged Irish and Jews, in contrast, show no difference in symptoms according to birth order.

On two successive interviews over a 2-year period this pattern for the disadvantaged and for the advantaged groups is repeated, with one exception. The exception is found among Irish respondents on the first interview, where none of the 9 first-born subjects reported four or more symptoms, while 5 of the 15 later-born subjects reported that many symptoms. This pattern is not stable, however, since, in the second interview with more respondents, 17% of the 12 first-born subjects, and 19% of the 16 later-born subjects reported four or more symptoms. In further analysis, therefore, we will continue to compare the Irish and Jews together, as relatively advantaged, with the disadvantaged Negroes and Puerto Ricans.

The implications of the finding that the general relation between advantaged or disadvantaged ethnic status, birth order, and psychological symptoms persists over two interviews, approximately 2 years apart, will be discussed after we consider possible alternative explanations of this relationship. First, the size of subjects' families of origin could be an artifact in the relationship shown in Table 3. Later-born subjects are more likely than first-born to come from large families, and members of the disadvantaged ethnic groups from large families might, in many cases, have suffered psychological damage as a consequence of economic hardship. As expected, many more of our later-born than first-born subjects came from families of five or more children, the median family size in this sample. Among subjects of disadvantaged ethnic status, 47% of first-born and 71% of later-born subjects came from families of five or more children. A higher rate of symptoms among subjects from large families could, therefore, explain the relation between birth order and symptoms among persons of disadvantaged ethnic status. As we see in Table 4, however, this explanation will not serve, since there tend to be proportionately more subjects with four or more symptoms from small than from large families.

If not large family of origin, then parental deprivation during childhood might explain the high level of symptoms among later-born subjects of disadvantaged ethnic status, since one or both parents are more likely to have died during the childhood of a later-born subject than of a first-born subject. This factor does not seem to explain the excess of symptoms among our later-born disadvantaged subjects, however, since the proportion of later-born disadvantaged subjects who lost one or both parents before the age of 12, 22%, is about the same as the proportion among first-born disadvantaged subjects, 29%. No meaningful comparison is possible with the advantaged subjects, since only five first-born and two later-born subjects lost one or both parents before the age of 12.

TABLE 4

RELATION OF NUMBER OF PSYCHOLOGICAL SYMPTOMS TO ETHNIC STATUS AND SIZE OF FAMILY OF ORIGIN AMONG LATER-BORN SUBJECTS

Number of symptoms (second interview)	Ethnic status			
	Disadvantaged (Negro and Puerto Rican)		Advantaged (Irish and Jewish)	
	Number of children in respondent's family of origin[a]			
	Less than five %	More than five %	Less than five %	More than five %
Less than 4	53	69	79	84
4 or more	47	31	21	16
Total	100	100	100	100
Number of subjects	15	36	24	19

Note.—In order to determine whether large family of origin is associated with four or more symptoms among later-born subjects of disadvantaged ethnic status, these subjects are compared with later-born advantaged subjects only; comparison with first-born disadvantaged subjects would not be useful because there are no subjects in this group with four or more symptoms.

[a] Median size of family of origin for all respondents in sample is five.

TABLE 5

RELATION OF NUMBER OF PSYCHOLOGICAL SYMPTOMS TO ETHNIC STATUS AND TO INCOME AND EDUCATION AMONG LATER-BORN SUBJECTS

Number of symptoms (second interview)	Ethnic status					
	Disadvantaged (Negro and Puerto Rican)			Advantaged (Irish and Jewish)		
	Income			Income		
	Less than $5000 %	$5000–$7499 %	$7500 or more %	Less than $5000 %	$5000–$7499 %	$7500 or more %
Less than 4	50	78	67	78	92	75
4 or more	50	22	33	22	8	25
Total	100	100	100	100	100	100
Number of subjects[a]	12	23	15	9	13	20
	Education			Education		
	Less than 8 years %	8–12 years %	13 or more %	Less than 8 years %	8–12 years %	13 or more years %
Less than 4	67	69	80	(1)	81	89
4 or more	33	31	20	—	19	11
Total	100	100	100		100	100
Number of subjects[a]	15	26	10	1	32	9

Note.—In order to determine whether income or education is associated with four or more symptoms among later-born subjects of disadvantaged ethnic status, these subjects are compared with later-born advantaged subjects only; comparison with first-born disadvantaged subjects would not be useful because there are no subjects in this group with four or more symptoms.

[a] Totals are less than those shown for later-born subjects in Table 1 due to loss of a few subjects who did not respond.

Another possibility to be considered is that the relation between social status, birth order, and symptomatology can be explained more simply in terms either of income or educational differences associated with both social status and symptomatology. Symptoms of psychological disorder are generally found to be inversely related to income and other closely related variables (Dohrenwend & Dohrenwend, 1965), and income and education are both disproportionately low in the disadvantaged ethnic groups. As Table 5 shows, however, neither income nor education explains the relatively large number of subjects with four or more symptoms among later-born disadvantaged subjects. At almost every level of income and education, the proportion of subjects with four or more symptoms remains about twice as high in the disadvantaged as in the advantaged groups. Only in the income category of $7,500 or more is the distribution of symptoms in the advantaged and disadvantaged groups approximately the same. However, this similarity does not explain the difference between first-born and later-born subjects in the disadvantaged groups, since this income level is more frequent among later-born than among first-born subjects. Only 18% of first-born subjects in the disadvantaged groups had a family income of $7,500 or better, while 30% of the later-born subjects in these groups were at this income level.

The important question of the interpretation of the persistence of the pattern of symptoms over a 2-year period commands further attention. This persistence could be explained in either of two ways. It may reflect personality defects in later-born subjects of disadvantaged ethnic status or, alternatively, it may indicate that conditions in which these subjects are able to cope effectively with anxiety are somewhat inaccessible to them. In the first case, the difference between first- and later-born subjects over time would be based on an underlying stability in individual symptomatology. The second explanation is indicated if we find that, despite the persistence of the group difference, individual symptom levels of later-born subjects of disadvantaged ethnic status vary over time.

The data in Table 6 support the second alternative. Thirty percent of later-born subjects of disadvantaged ethnic status moved into or out of the group having four or more symptoms during the 2 years intervening between interviews, the number moving in each direction being about equal. Only 16% of these subjects report four or more symptoms on both interviews. Thus, there appears to be little support for the position that a high level of psychological symptomatology in later-born persons reflects persistent personality defects in this group. A third measure taken at a later time might, of course, show that there are two types of subjects, those whose symptoms fluctuate and those whose symptoms are stable at a high or low level. It might, for instance, be found that the 16% of later-born subjects of disadvantaged status who reported four or more symptoms on these two interviews continue to report as many symptoms. On the basis of the present evidence, however, there is no reason to assume such stability, particularly in view of the fact that only about a quarter of the subjects of disadvantaged status show no change in symptom level at all between the two interviews.

While the absence of stability in individual symptomatology is apparent, these changes still require an explanation. The second interview did not include a systematic survey of stressors impinging on the subjects, but questions concerning illness and injury in the family, births, and job changes yielded some relevant information. When responses to these questions were classified according to whether they indicated either favorable or unfavorable changes currently impinging on the subject, it was found that 56 subjects could be identified as currently influenced either by entirely favorable or entirely unfavorable events. The directions of the symptom changes in these subjects, as well as in subjects not so classified, are shown in Table 7. Unfortunately, there are too few cases to permit simultaneous control of birth order and ethnic status.

Table 7 shows that the relation between symptom changes and the quality of currently impinging events is significant. It also presents the categories of favorable and unfavorable events to show that the results are not due solely to changes in physical health, with which the 22-item symptom measure may be contaminated, or to socioeconomic changes which could be consequences rather than causes of the symptoms in the main earner. It shows, for example, that symptoms

TABLE 6

RELATION OF ETHNIC STATUS TO CHANGES IN SYMPTOMATOLOGY BETWEEN FIRST AND SECOND INTERVIEWS

Number of symptoms at		Ethnic status			
		Disadvantaged		Advantaged	
		First born %	Later born %	First born %	Later born %
First interview	Second interview				
Less than four	Less than four	94	54	75	70
Four or more	Four or more	0	16	7	15
Less than four	Four or more	0	16	11	5
Four or more	Less than four	6	14	7	10
Total		100	100	100	100
Number of subjects		17	49	27	40
Direction of change at all levels of symptomatology		%	%	%	%
No change		29	22	26	50
Fewer symptoms at 2nd interview		42	35	30	32
More symptoms at 2nd interview		29	43	44	18
Total		100	100	100	100
Number of subjects		17	49	27	40

TABLE 7

DIRECTION OF SYMPTOM CHANGE IN RELATION TO QUALITY OF EVENTS CURRENTLY IMPINGING ON SUBJECTS AT TIME OF SECOND INTERVIEW

Quality of events	Direction of symptom change				Number of subjects
	Fewer symptoms on second interview	More symptoms on second interview	No change		
			Same number of symptoms on both interviews	No symptoms on either interview	
Favorable—% totals 100	41%	23%	13%	23%	31
Socioeconomic improvement					
Main male earner—*n*	5	5	2	4	16
Wife of main earner—*n*	4	2	2	1	9
Birth of children					
Wife—*n*	2	—	—	—	2
Husband—*n*	—	—	—	2	2
Combinations of above—*n*	2	—	—	—	2
Unfavorable—% totals 100	16%	56%	20%	8%	25
Poor physical health of					
Self—*n*	—	7	2	—	9
Spouse—*n*	2	3	3	2	10
Socioeconomic deterioration					
Main male earner—*n*	1	2	—	—	3
Wife of main earner—*n*	—	2	—	—	2
Combinations of above—*n*	1	—	—	—	1
Mixed[a] or unknown—% totals 100	37%	31%	17%	15%	80
$\chi^2 = 5.39^b$ $p < .05$					

[a] Four cases.

[b] Favorable versus Unfavorable Events × Fewer versus more Symptoms on Second Interview.

tend to increase for respondents whose spouses become ill, and that wives' symptoms tend to increase or decrease with negative or positive changes respectively in their husbands' job situations. These results suggest, therefore, the importance of further investigation of subjects' contemporary situation in attempts to explain psychological symptomatology (cf. Haberman, 1965).

Turning again to the relatively high rate of symptoms among later-born subjects of the lead suggested above that these symptoms are reactions to contemporary events, the investigators attempted to find conditions in the subjects' current situations which would explain their symptoms in terms of the adaptive mechanisms described by Schachter (1959) and by Wrightsman (1960). Although a number of conditions were investigated which appear characteristic of the lives of the disadvantaged ethnic groups and would tend to enforce social interaction and prevent isolation, none were found which account for the higher symptom level among later-born members of the disadvantaged ethnic groups. Specifically, neither person-per-room ratio nor absolute number of persons in a living unit is higher for later-born subjects with high symptoms than for others. Nor does gross division of the area studied into a predominantly slum and a predominantly nonslum section show a lower rate of symptoms among the relatively few later-born members of the disadvantaged groups in the nonslum area. This division is not perfect, however, since there are pockets of relatively good buildings in the predominantly slum areas, and pockets of slum buildings in the predominantly nonslum area.

DISCUSSION

While the affiliative reactions of first- and later-born subjects in anxiety-arousing conditions help to account for the number of psychological symptoms they experience in experimental social isolation, the evidence is less decisive on the part these mechanisms play in determining the striking differences found in the community. Heretofore, such differences in community and patient groups have been interpreted as evidence of persistent personality defect in the sibling type with the higher rate of symptoms. The large amount of fluctuation in the level of symptoms in our community sample over time, however, is indirect evidence for the alternative hypothesis which we derived from Schachter's work: symptom levels of first- and later-born persons are a function of the extent to which social factors in contemporary life situations facilitate or block the type of affiliative response which enables them to reduce anxiety most effectively; specifically, first borns show high levels of symptoms when their situation is such that, faced with an anxiety-arousing condition, they have difficulty finding others with whom to share the experience. Later borns show high levels of symptoms when, in the face of anxiety, they are forced into intense interaction with others about the relevant experience. Direct tests of these hypotheses will require more intensive research on first- and later-borns' patterns of social interaction with family, friends, neighbors, and others in the community.

Such research would constitute a further step toward resolution of the issue which has emerged in the recent literature on birth order, that is, whether observed differences in levels of psychological distress of first- and later-born persons are a function of their contemporary social situation, or of basic personality defect in one or the other group. Studies of a variety of indicators of psychological distress over time would also be necessary to determine whether such indicators are stable at the individual level. For those symptoms found to be unstable, the hypotheses above could be applied directly to explain differences between first and later borns. For instances where the evidence over time supports the interpretation of stable differences in rate of persistent personality defect between first and later borns, other mechanisms may be involved. To date, however, convincing evidence over time of the existence of the latter differences has not been presented.

REFERENCES

ALTUS, W. D. Birth order, intelligence, and adjustment. *Psychological Reports,* 1959, **85**, 502.

DOHRENWEND, B. P., & DOHRENWEND, B. S. The problem of validity in field studies of psychological disorder. *Journal of Abnormal Psychology,* 1965, **70**, 52–69.

ELINSON, J., & LOWENSTEIN, R. *Community fact book for Washington Heights, New York City, 1960–1961.* New York: Columbia University School of Public Health and Administrative Medicine, 1963.

GREENBERG, H., GUERINO, R., LASHEN, M., MEYER, D., & PISKOWSKI, D. Order of birth as a determinant of personality and attitudinal characteristics. *Journal of Social Psychology,* 1963, **60**, 221–230.

HABERMAN, P. W. An analysis of re-test scores for an index of psycho-physiological disturbance. *Journal of Health and Human Behavior,* 1965, **6**, 257–260.

LANGNER, T. S. A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health and Human Behavior,* 1962, **3**, 269–276.

PRAKASH, J., & SRIVASTAVA, P. K. The effect of family size, economic condition and order of birth on adolescent adjustment. *Manas,* 1963, **10**, 83–88.

SARNOFF, I., & ZIMBARDO, P. G. Anxiety, fear, and social affiliation. *Journal of Abnormal and Social Psychology,* 1961, **62**, 356–363.

SCHACHTER, S. *The psychology of affiliation.* Stanford, Calif.: Stanford University Press, 1959.

SCHOOLER, C. Birth order and schizophrenia. *Archives of General Psychiatry,* 1961, **4**, 91–97.

VOGEL, W., & LAUTERBACH, C. G. Sibling patterns and social adjustment among normal and psychiatrically disturbed soldiers. *Journal of Consulting Psychology,* 1963, **27**, 236–242.

VUYK, R. Eltern vergleichen ihre beiden Kinder zum zweiten Mal. *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen,* 1963, **22**, 220–231.

WILSON, R. S. On behavior pathology. *Psychological Bulletin,* 1963, **60**, 130–146.

WRIGHTSMAN, L. S., JR. Effects of waiting with others on changes in level of felt anxiety. *Journal of Abnormal and Social Psychology,* 1960, **61**, 216–222.

(Received June 8, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 224–231

RIGIDITY IN THE RETARDED:

A FURTHER TEST OF THE LEWIN-KOUNIN FORMULATION[1]

EDWARD ZIGLER AND EARL C. BUTTERFIELD

Yale University

Groups of younger and older familial retardates from 2 institutions and a group of normals were compared on 3 of Kounin's measures of rigidity. The general findings were inconsistent with the Lewin-Kounin rigidity formulation and the results reported by Kounin. The differences obtained between the groups investigated here are discussed in terms of a number of cognitive and motivational factors which have been advanced to explain differences in performance between normals and retardates of the same MA.

Few formulations have had greater impact on the area of mental retardation than the rigidity hypothesis of Lewin (1936) and Kounin (1941). Lewin argued that, as compared to normal children of the same degree of differentiation, that is, MA, retardates were characterized by a lessened permeability in the boundaries between cognitive regions. This position was extended somewhat by Kounin who argued that the degree of rigidity in these boundaries was a positive monotonic function of CA. This argument generated the prediction that, with MA held constant, older retardates suffered from greater cognitive rigidity than younger retardates who in turn were more rigid than normal children. This prediction was confirmed in a series of five experiments conducted by Kounin. Kounin's findings have probably stimulated more research (e.g., Plenderleith, 1956; Shallenberger & Zigler, 1961; Stevenson & Zigler, 1957; Thompson & Magaret, 1947; Zigler, 1961, 1963; Zigler & DeLabry, 1962; Zigler, Hodgden, & Stevenson, 1958) and theoretical debate (Goldstein, 1942; Kounin, 1948; Sarason & Gladwin, 1958; Werner, 1948; Zigler, 1962) than any other single experimental study in the psychological literature on mental retardation. Furthermore, as Sarason and Gladwin (1958) have noted, the Lewin-Kounin view that the retarded are rigid has had considerable influence on the care, training, and treatment of the retarded.

The work of Zigler and his colleagues (reviewed in Zigler, 1962, 1966) has now indicated that much of the rigid behavior of retardates can be attributed to a variety of motivational factors, rather than to an inherent cognitive rigidity. However, with the exception of Kounin's concept-switching task (Zigler & DeLabry, 1962; Zigler & Unell, 1962), none of the studies which have cast doubt upon the Lewin-Kounin formulation have employed other actual experimental tasks used by Kounin. In view of the evidence that has now been presented, that familial retardates are no more rigid than normals of the same MA, as well as the theoretical and practical significance of Kounin's work, a further examination of the performance of retardates on Kounin's tasks appears appropriate.

Kounin's findings were obtained with retarded subjects run within a single institution. It is now apparent that institutionalization cannot be treated as a homogeneous psychological variable (Yarrow, 1964; Zigler, 1964). In respect to the retarded, evidence has now been presented (Butterfield & Zigler, 1965) that differing institutional milieux affect behaviors on a task in which the child complies with the instructions of an adult. This instruction-initiated feature is characteristic of the tasks employed by Kounin. In light of this, and in order to assess the generality of the present findings, this study em-

---

[1] This research was supported by Research Grant MH-06809 from the National Institute of Mental Health, United States Public Health Service, and the Gunnar Dybwad Award of the National Association for Retarded Children. The authors wish to thank the superintendents and staff members of Institutions A and B and members of the New Haven, Connecticut public school system for their cooperation.

TABLE 1

MA AND CA FOR EACH OF THE FIVE GROUPS IN THE PRESENT INVESTIGATION AND FOR EACH OF KOUNIN'S (1941) THREE GROUPS

	N	MA			CA		
		$\bar{X}$	Range	SD	$\bar{X}$	Range	SD
Present investigation							
Institution A							
Young	20	86.0	72–95	6.3	187.0	95–205	14.7
Old	20	85.0	72–93	6.9	483.0	365–580	71.8
Institution B							
Young	20	84.0	74–96	7.2	188.0	143–211	24.5
Old	20	85.0	72–96	7.0	485.0	372–571	63.4
Normal	20	84.7	72–96	7.1	75.8	64–86	6.4
Kounin's investigation							
Young	21	82.0	72–96	—[a]	174.0	130–204	—[a]
Old	21	80.0	69–94	—[a]	500.0	351–645	—[a]
Normal	21	82.0	72–92	—[a]	80.0	72–93	—[a]

[a] Not reported by Kounin.

ployed two institutions for the retarded which had distinctly different social climates.

METHOD

Subjects

Four groups of 20 retarded subjects equated on WISC or Stanford-Binet MAs were selected. One group of chronologically young and one group of older retardates were selected from each of two institutions. All of the retarded subjects were of the familial type, and no subjects with obvious sensory or motor impairments were used. A fifth group, composed of 20 normal children from a New Haven public school, was employed. Prior work in the New Haven school system indicated a mean IQ of 112, and this IQ was employed in assessing the mean MA of the normal group. One half of the subjects in each group were male, and one half were female. The characteristics of the five groups are presented in Table 1.

Institutions

The two institutions were those employed by Butterfield and Zigler (1965) who have described these two institutions at length. In summary, Institution A has been characterized as an enlightened setting in which the climate is conducive to constructive, supportive interactions between the residents and their caretakers. In contrast, Institution B has been characterized as being restrictive, having a custodial philosophy, and one in which there are limited interactions between residents and caretakers.

Experimental Tasks

Three of the five tasks employed by Kounin (1941) were used. The particular tasks chosen were the ones which the authors felt provided the clearest test of the Lewin-Kounin rigidity hypothesis (see Kounin, 1941, for a more detailed description of these three tasks).

Transfer of habit. During a training period, the experimenter inserted individual marbles into the top of a small box which the subject retrieved by *depressing* a lever on the front of the box. After 90 training trials, the subject was instructed to secure the marble by *raising* the lever. The measure employed was the number of lever depressions in the 60 trials following the instructions to raise the lever.

Satiation and cosatiation. The subject was trained in the drawing of four simple line drawings: cats, bugs, turtles, and rabbits. He was then instructed to draw cats until he wished to draw no more of them. The subject was then asked if he wanted to draw bugs and, if he did, he was allowed to draw bugs until he wished to stop. He was then asked if he wished to draw turtles, and then rabbits. The situation was made as free as possible and there was a minimum of pressure to continue drawing. The criterion measures derived from this task were:

$$\frac{\text{time spent drawing cats} - \text{time spent drawing bugs}}{\text{time spent drawing cats}} \times 100 \qquad [1]$$

$$\frac{\text{cats} - \text{turtles}}{\text{cats}} \times 100 \qquad [2]$$

$$\frac{\text{cats} - \text{rabbits}}{\text{cats}} \times 100 \qquad [3]$$

$$\frac{\text{cats} - \dfrac{(\text{bugs} + \text{turtles} + \text{rabbits})}{3}}{\text{cats}} \times 100 \qquad [4]$$

Restructuring by classification. Each subject demonstrated that he was able to sort cards which differed in color only (black, yellow, green, pink, and red) and cards which differed in form only (triangle, square, circle, cross, and five-pointed star). The subject was then asked to sort a deck of 25 cards which differed in both these colors and forms (five triangles, one of each color; five squares, one of each color, etc.). All subjects successfully sorted this mixed deck on the basis of either color or form. The subject was then asked to sort this deck again. If he did not change his basis for classification (from shape to color or from color to shape) he was asked to sort them again. If he still had not changed his classification scheme he was told to "try to find a different way they're alike," and given six more opportunities to sort the cards. The measure employed was the trial upon which the subject switched his classification principle. If he did not switch he was given a score of 10.

All subjects were seen individually by a young adult male who was unfamiliar to them. He was pleasant but businesslike throughout his contact with the subjects. The order of criterion task administration was randomly determined for each subject. All tasks were administered in a single session that averaged about 1 hour in length.

RESULTS

Preliminary analyses revealed no significant sex effects for any of the tasks, and for purposes of clarity the sex factor is not included in the following analyses.

Transfer of Habit

The Lewin-Kounin hypothesis that the boundaries between cognitive zones were most permeable in normal children, less permeable in retarded children, and least permeable in retarded adults generated the prediction that these groups would evidence differing amounts of negative transfer. The specific prediction that the normal children would make the most errors (that is, depressing the lever when they should be raising it), retarded children an intermediate number of errors, and retarded adults the least errors was confirmed by Kounin (see Table 2). No support for this prediction was found in the present study. A simple, randomized analysis of variance revealed no significant differences among the five groups in number of lever depressions $(F_{4/95} = 1.30, p > .75)$.

The effects of age and institution upon the performances of the four retarded groups were examined by means of a 2 × 2 (Age × Institution) factorial analysis of variance. There was a trend ($F = 3.28$, $.10 > p > .05$) for the older retardates to make more errors

TABLE 2

CRITERION SCORES FOR THE FIVE EXPERIMENTAL GROUPS AND FOR KOUNIN'S (1941) THREE EXPERIMENTAL GROUPS

	Transfer of habit errors		Cosatiation in percentage								Restructuring by classification	
			$\frac{\text{cats} - \text{bugs}}{\text{cats}}$		$\frac{\text{cats} - \text{turtles}}{\text{cats}}$		$\frac{\text{cats} - \text{rabbits}}{\text{cats}}$		$\frac{\text{cats} - \frac{(\text{bugs} + \text{turtles} + \text{rabbits})}{3}}{\text{cats}}$			
	X	*SD*	*X*	*SD*	*X*	*SD*	*X*	*SD*	*X*	*SD*	No. switching	No. not switching
Present investigation												
Institution A												
Young	4.80	10.9	42	55	43	75	51	76	45	63	10	10
Old	11.70	21.2	19	69	37	56	44	56	35	54	2	18
Institution B												
Young	2.40	3.7	27	54	51	53	59	56	46	44	14	6
Old	7.25	16.0	41	50	39	77	47	60	47	58	4	16
Normal	6.00	8.3	12	49	−15	73	4	55	6	41	14	6
Kounin's investigation												
Young	2.85	.97	39	—[a]	52	—[a]	66	—[a]	54	—[a]	12	8
Old	1.20	.47	−28	—[a]	− 8	—[a]	8	—[a]	−10	—[a]	6	14
Normal	7.00	.99	90	—[a]	90	—[a]	96	—[a]	92	—[a]	21	0

[a] Not reported by Kounin.

than the young retardates. This trend is in the opposite direction to the prediction and findings of Kounin. The main effect for institutions ($F = 1.10$) and the Age × Institution interaction ($F < 1.0$) did not approach significance. As can be seen in Table 2, the standard deviations obtained in this study were much larger than those obtained by Kounin.

Satiation and Cosatiation

The rigidity-of-boundaries position generated the prediction that old retardates would show less cosatiation than young retardates who, in turn, would show less cosatiation than normal children. The view here is that the thicker the boundary the less should satiation in one region affect a neighboring region. Kounin's results clearly supported this hypothesis (see Table 2).

Each of the four cosatiation scores obtained by each of the five groups in the present study were compared by means of simple randomized analyses of variance (see Table 2). There were no significant differences ($F_{4/95} = 1.06$) on the first score, that is,

$$\frac{\text{time spent on cats} - \text{time spent on bugs}}{\text{time spent on cats}} \times 100.$$

A significant difference ($F_{4/95} = 3.03, p < .05$) between groups was found on the second cosatiation score, that is, $\frac{\text{cats} - \text{turtles}}{\text{cats}} \times 100.$ Analysis by t tests indicated that all retarded groups showed more cosatiation than the average group ($p < .05$ in each case) with no significant differences being found between any of the retarded groups. The third, that is, $\frac{\text{cats} - \text{rabbits}}{\text{cats}} \times 100$ and fourth cosatiation scores, that is,

$$\frac{\text{cats} - \dfrac{(\text{bugs} + \text{turtles} + \text{rabbits})}{3}}{\text{cats}} \times 100,$$

both revealed trends ($p < .10$) toward differences between the groups ($F_{4/95} = 2.42$, and 2.28, respectively). The trends were for each of the retarded groups to have higher cosatiation scores than the normal group. The one significant difference and the trends found were in the direction opposite to that predicted by Kounin.

These differences in cosatiation scores primarily reflected the higher frequency with which retarded subjects, regardless of their age or institution, declined to draw bugs, turtles, and/or rabbits. The frequency of declining to draw in each of the five groups is shown in Table 3, which also presents chi-square comparisons of frequency of declining between every possible combination of two of the five groups. Every retarded group declined more frequently than the average group, and there were no significant differences in frequency of declining among the retarded groups. A comparison of the combined young and old retardate groups from both institutions A and B showed that young retardates declined more frequently than old retardates ($\chi^2 = 4.06$, $p < .05$). (Analyses, ignoring the institutional dimension, were run on the cosatiation scores of only those subjects who never declined and showed no significant differences between the three groups: largest $F_{2/57} = 1.37$.)

The effects of age and institution upon each of the retardate groups' four cosatiation scores were examined by means of 2×2 (Age × Institution) factorial analyses of variance. The only F which was greater than 1.0 in these four analyses (Age × Institution interaction for the first cosatiation score) was not statistically significant ($F_{1/76} = 1.94$).

Restructuring by Classification

On this task, in which the subject must move from one cognitive zone to another on his own, the Lewin-Kounin formulation generates the prediction that such movement would be most difficult for the old retardate and least difficult for the normal child. Analysis of Kounin's data (see Table 2) by the present authors showed that, as predicted, he found less switching from one concept to another among young retardates than normal subjects ($\chi^2 = 8.04$, $p < .01$), less switching among old retardates than normal subjects ($\chi^2 = 19.32$, $p < .001$), and a difference of borderline significance in frequency of shifting between young and old retardates ($\chi^2 = 2.52$, $.10 < p < .20$).

TABLE 3

FREQUENCY OF DECLINING AND NOT DECLINING TO DRAW TURTLES, BUGS, AND/OR RABBITS IN EACH OF THE FIVE GROUPS AND CHI-SQUARE VALUES ASSOCIATED WITH EVERY POSSIBLE COMPARISON

	Frequencies		Chi-square values				
			Institution A		Institution B		
	Declining	Not declining	Young	Old	Young	Old	Normal
Institution A							
Young	13	7	—	2.50	<1.0	2.50	13.29***
Old	7	13		—	<1.0	1.0	3.90*
Institution B							
Young	11	9			—	<1.0	9.64**
Old	7	13				—	3.90*
Normal	1	19					—

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Chi-squares on frequency of switching were computed for all possible combinations of the five groups in the present study, and are presented in Table 4. They show that the two young retarded groups and the normal group were not significantly different from one another in frequency of switching, but that all three of these groups switched more frequently than did both old retarded groups. The two old retarded groups were not significantly different from one another. By combining the young and old groups in Institutions A and B it was found that there was significantly more switching at Institution B than at Institution A (Fisher's exact test; $p = .032$). A graphic presentation of Kounin's and the present study's switching data are presented in Figure 1.

DISCUSSION

The general findings of the present study lend little support to the formulation that familial retardates are inherently more rigid than normals of the same MA. No significant differences were found between groups on the transfer-of-habit task. A trend of border-

TABLE 4

FREQUENCY OF SWITCHING AND NOT SWITCHING IN EACH OF THE FIVE GROUPS AND CHI-SQUARE VALUES ASSOCIATED WITH EVERY POSSIBLE COMPARISON

	Frequency		Chi-square values				
			Institution A		Institution B		
	Switching	Not switching	Young	Old	Young	Old	Normal
Institution A							
Young	10	10	—	5.83*	.93	2.74	.93
Old	2	18		—	12.60***	.19	12.60**
Institution B							
Young	14	6			—	8.18**	0
Old	4	16				—	8.18**
Normal	14	6					—

* $p < .02$.
** $p < .01$.
*** $p < .001$.

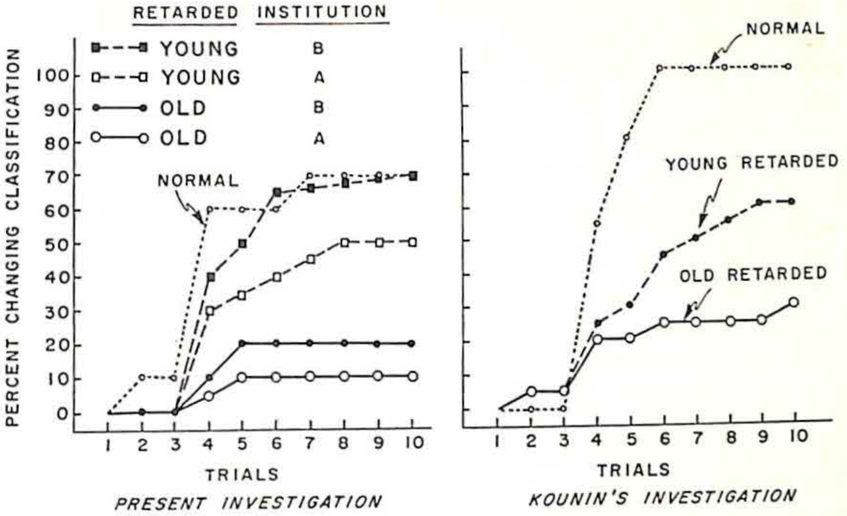


Fig. 1. Cumulative percentage by trials of subjects changing sorting classification in the present and in Kounin's investigations.

line significance was found indicating that older retardates made more errors than younger retardates, and were thus less rigid. This trend is in opposition to the rigidity formulation and the findings of Kounin.

On the drawing task, the one significant finding and the two trends found in the overall comparisons indicated higher cosatiation scores in the normal than in the four retarded groups. This is also in direct opposition to the rigidity hypothesis and Kounin's findings. No one would argue that this finding indicates that normals are more rigid than retardates. Rather, it would appear to call into question the validity of Kounin's cosatiation task as a measure of cognitive rigidity. The inadequacy of this measure is highlighted by the fact that it is the tendency of retardates simply to decline to draw which results in their higher cosatiation scores. The strong tendency of retardates to decline to draw is consistent with the anecdotal and experimental evidence that retardates suffer from a reluctance or wariness to interact with strange adults (Hirsh, 1959; Sarason & Gladwin, 1958; Wellman, 1938; Woodward, 1960), a negative-reaction tendency (Shallenberger & Zigler, 1961; Zigler, 1966), and/or a general avoidance of situations in which they might experience failure (Cromwell, 1963).

The fact that older retardates declined to draw less often than younger retardates is open to a number of interpretations: (*a*) older retardates who are not involved in academic programs may not see drawing as involving intellectual evaluation; (*b*) lengthened institutionalization has been shown to result in a heightened motivation to interact with an attentive adult (Butterfield & Zigler, 1965; Zigler & Williams, 1963) which could attenuate the tendency to decline; (*c*) lengthened institutionalization could result in a general increase in compliance making the subject feel less free to decline; and, (*d*) lengthy residence in a protective institutional environment may reduce the intensity of the failure-avoiding motive, making the retardate more willing to engage in tasks which he might conceptualize as demanding. These myriad possibilities, as well as the general findings obtained with the cosatiation index in other studies (McCoy & Zigler, 1965; Shallenberger & Zigler, 1961; Zigler et al., 1958), strongly suggest that this score, rather than being a measure of cognitive rigidity, reflects a variety of motivational factors.

The finding that young retardates do not differ significantly from normals on the concept-switching task is inconsistent with Kounin's findings, as well as certain of those of Zigler and DeLabry (1962), but is in keeping with those of Zigler and Unell (1962). The finding that older retardates had more trouble switching concepts than did the other groups is the one result of this study consistent with Kounin's formulation. In light of the tendency of the older retardates to manifest the least rigidity on the transfer-of-habit task, it is difficult to interpret their performance on the concept-switching task as clearly indicating greater rigidity. An alternative to the rigidity formulation may be found in the study of Iscoe and McCann (1965) who presented evidence that long-time institutionalization results in a generally reduced ability to form concepts. Mitigating against this possibility is the fact that the older retardates had no difficulty in forming color and form concepts, but only in switching from one to the other. Adding to the complexity here are the findings that whether differences between normals and retardates are obtained on this particular task depends upon the social class of the normals and the particular reward being employed to motivate the subject to switch concepts (Zigler & DeLabry, 1962; Zigler & Unell, 1962). It thus appears that performance on Kounin's concept-switching task is sensitive to other than purely cognitive factors. This can be seen in the present study in the finding that retardates in the poorer institution switched more readily than subjects in the better institution, a finding that casts further doubt on Kounin's general rigidity formulation. It hardly seems logical to argue that the better social climate makes for greater cognitive rigidity or poorer conceptual ability. The institutional difference discovered must be mediated by some as yet undefined motivational factor. Clarification of this issue must await further studies of the older retardate as well as the study of motivational and cognitive effects of long-term institutionalization in institutions having different psychosocial environments.

Certain procedural differences in the selection of subjects could account for the discrepancies between Kounin's and the present findings. Kounin obtained normal subjects by simply asking for volunteers from a classroom. He included retarded subjects only after a careful screening procedure in which they affirmed their desire to participate. Over half of Kounin's original retarded subjects were excluded as a result of this procedure. Furthermore, Kounin spent a week with certain of his retarded subjects in an attempt to develop friendly relationships with them before beginning his experiment. In the present investigation, both normal and retarded subjects were selected randomly and were given no opportunity to decline to participate and no subjects had any pre-experimental contact with the experimenter. That Kounin's selection procedures may have resulted in his group being more homogeneous is suggested by the much smaller standard deviations of his groups in the transfer-of-habit task (Table 2). Although Kounin does not feel that these are the critical factors,[2] it does seem possible that they could have been instrumental in producing the differences between the results of the earlier and the present study.

Since findings inconsistent with those of Kounin could not have been confidently predicted, the present study did not test for the importance of any of these specific variables in producing the differences between Kounin's and the present study. However, several potentially fruitful lines of investigation are suggested. Among these would be the significance of institutionalization versus non-institutionalization, length of institutionalization, nature of the institutional setting, subject variables which differentiate the volunteering from the nonvolunteering subject, and the nature of the preexperimental interaction between the subject and the experimenter. The present study does highlight the need for further investigation of those alternatives to the Lewin-Kounin interpretation of the rigid behaviors of retardates that have been advanced by the present investigators (Butterfield & Zigler, 1965; Zigler, 1962, 1966) and other workers as well (Luria, 1963; Siegel & Foshee, 1960; Spitz, 1963).

---

[2] Jacob S. Kounin, Personal communication, 1964.

REFERENCES

BUTTERFIELD, E. C., & ZIGLER, E. The influence of differing institutional social climates on the effectiveness of social reinforcement in the mentally retarded. *American Journal of Mental Deficiency*, 1965, **70**, 48–56.

CROMWELL, R. L. A social learning approach to mental retardation. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963. Pp. 41–91.

GOLDSTEIN, K. Concerning rigidity. *Character and Personality*, 1942, **11**, 209–226.

HIRSH, E. A. The adaptive significance of commonly described behavior of the mentally retarded. *American Journal of Mental Deficiency*, 1959, **63**, 639–646.

ISCOE, I., & MCCANN, B. Perception of an emotional continuum by older and younger mental retardates. *Journal of Personality and Social Psychology*, 1965, **1**, 383–385.

KOUNIN, J. Experimental studies of rigidity. I. The measurement of rigidity in normal and feebleminded persons. *Character and Personality*, 1941, **9**, 251–273.

KOUNIN, J. The meaning of rigidity: A reply to Heinz Werner. *Psychological Review*, 1948, **55**, 157–166.

LEWIN, K. A dynamic theory of personality. New York: McGraw-Hill, 1936.

LURIA, A. R. Psychological studies of mental deficiency in the Soviet Union. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963. Pp. 353–387.

MCCOY, N., & ZIGLER, E. Social reinforcer effectiveness as a function of the relationship between child and adult. *Journal of Personality and Social Psychology*, 1965, **1**, 604–612.

PLENDERLEITH, M. Discrimination learning and discrimination reversal learning in normal and feebleminded children. *Journal of Genetic Psychology*, 1956, **88**, 107–112.

SARASON, S. B., & GLADWIN, T. Psychological and cultural problems in mental subnormality: A review of research. *Genetic Psychology Monographs*, 1958, **57**, 3–289.

SHALLENBERGER, P., & ZIGLER, E. Rigidity, negative reaction tendencies, and cosatiation effects in normal and feebleminded children. *Journal of Abnormal and Social Psychology*, 1961, **63**, 20–26.

SIEGEL, P. S., & FOSHEE, J. G. Molar variability in the mentally defective. *Journal of Abnormal and Social Psychology*, 1960, **61**, 141–143.

SPITZ, H. H. Field theory in mental deficiency. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill. Pp. 11–40.

STEVENSON, H., & ZIGLER, E. Discrimination learning and rigidity in normal and feebleminded individuals. *Journal of Personality*, 1957, **25**, 699–711.

THOMPSON, C., & MAGARET, A. Differential test responses of normals and mental defectives. *Journal of Abnormal and Social Psychology*, 1947, **42**, 285–293.

WELLMAN, B. Guiding mental development. *Childhood Education*, 1938, **15**, 108–112.

WERNER, H. The concept of rigidity: A critical evaluation. *Psychological Review*, 1948, **53**, 43–53.

WOODWARD, M. Early experiences and later social responses of severely subnormal children. *British Journal of Medical Psychology*, 1960, **33**, 123–132.

YARROW, L. J. Separation from parents during early childhood. In M. L. Hoffman & L. Hoffman (Eds.), *Review of child development research*. Vol. I. New York: Russell Sage Foundation, 1964. Pp. 89–136.

ZIGLER, E. Social deprivation and rigidity in the performance of feebleminded children. *Journal of Abnormal and Social Psychology*, 1961, **62**, 413–421.

ZIGLER, E. Rigidity in the feebleminded. In E. P. Trapp & P. Himelstein (Eds.), *Research readings on the exceptional child*. New York: Appleton-Century-Crofts, 1962. Pp. 141–162.

ZIGLER, E. Rigidity and social reinforcement effects in the performance of institutionalized and noninstitutionalized normal and retarded children. *Journal of Personality*, 1963, **31**, 258–269.

ZIGLER, E. The effect of social reinforcement on normal and socially deprived children. *Journal of Genetic Psychology*, 1964, **104**, 235–242.

ZIGLER, E. Research on personality structure in the retardate. In N. R. Ellis (Ed.), *International review of research in mental retardation*. Vol. I. New York: Academic Press, 1966. Pp. 77–105.

ZIGLER, E., & DELABRY, J. Concept-switching in middle class, lower class, and retarded children. *Journal of Abnormal and Social Psychology*, 1962, **65**, 267–273.

ZIGLER, E., HODGDEN, L., & STEVENSON, H. The effect of support on the performance of normal and feebleminded children. *Journal of Personality*, 1958, **26**, 106–122.

ZIGLER, E., & UNELL, E. Concept-switching in normal and feebleminded children as a function of reinforcement. *American Journal of Mental Deficiency*, 1962, **66**, 651–657.

ZIGLER, E., & WILLIAMS, J. Institutionalization and the effectiveness of social reinforcement: A three-year follow-up study. *Journal of Abnormal and Social Psychology*, 1963, **66**, 197–205.

(Received June 17, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 3, 232–234

SEXUALLY DEVIANT BEHAVIOR IN EXPECTANT FATHERS

A. ARTHUR HARTMAN AND ROBERT C. NICOLAY

The Psychiatric Institute Municipal Department, Circuit Court of Chicago

Loyola University

A group of 91 expectant fathers examined in a court psychiatric clinic is compared with a group of 91 nonexpectant married males equated for age, race, and year of arrest. The hypothesis that sexually deviant reactions are more frequent in expectant fathers is statistically substantiated. Other findings are reported and discussed including the frequent occurrence of sexual offenses during the wife's 1st pregnancy.

Pregnancy has long been recognized as a stress period for the mother. Symptoms commonly reported include emotional lability, mild depressions (maternity blues), flightiness, hostility toward the husband, and various forms of regressive behavior. In approximately one out of every 400 pregnancies and childbirths a psychotic reaction occurs. It is estimated that 10–15% of these psychotic reactions occur before the child is born (Weiner & Steinhilber, 1961). In a study of 100 puerperal psychoses, White, Prout, Fixsen, and Fondeur (1957) allude to immature, neurotic reactions on the part of the husband to his wife's pregnancy and his anticipated fatherhood, emphasizing as highly significant that these reactions may markedly augment the stressfulness of the entire situation for the wife.

The purpose of the present study was to investigate the association between pregnancy and antisocial reactions manifested by expectant fathers. These were persons who had been arrested for various offenses and referred by judges to the psychiatric clinic for examination.

A search of the literature revealed four studies reporting data on stress reactions of expectant fathers. The earliest of the studies (Zilboorg, 1931) focused on depressive reactions of 30 patients. Noting unconscious hatred of the father and rivalry with the new child, Zilboorg postulated fear of incest as basic to the fear of parenthood. Following in the same vein of psychoanalytic theory, but attempting to explain the symptom choice, Freeman (1951) reported six cases of mental illness, the onset of which coincided with the pregnancy of the patients' wives. The wife's pregnancy was viewed as a precipitating factor in predisposed persons. Unable to find suitable outlets for extremely intense impulses of aggression, fear, and rivalry, they retreat into symptom formation to reduce the intrapsychic tension.

Towne and Afterman (1955), reported a study of 28 expectant fathers hospitalized in a Veterans Administration psychiatric setting. Symptoms of hostility and rivalry were reported but in contrast to findings of other investigations dependency needs were seen as more primary and ubiquitous than incest conflict.

In the only study employing comparison groups, Curtis (1955) explored factors differentiating the emotional adaptation of three groups to fatherhood. From an analysis of projective test data of 55 expectant fathers, these groups were differentiated according to "unconscious" images of themselves. Those referred for nonpsychiatric reasons maintained a relatively stable "unconscious" image of themselves as capable, loving father figures. Men referred for psychiatric examination with "minor" problems less often achieved the "good father" identification but were able to experience a comfortable unconscious identification as a "good" mother or older sibling. Men referred for "serious" problems were unable to make any comfortable, stable identification.

While the literature cited documents the concept of pregnancy as a stress process upon the expectant father, the explanations offered are largely impressionistic, sketchy, and embarrassed by the limited empirical data. The present study offers data to assist in mapping this largely unexplored area.

The frequency of sexually deviant behavior among arrested men whose wives were pregnant was a matter of comment by the staff of a court psychiatric clinic. This sexual behavior included exhibitionism, pedophilia, rape, homosexual acts, incest, transvestitism, obscene phone calls and letter writing, and public masturbation. A systematic collection of data was begun to test the following hypotheses:

1. Sexual offenses are more frequent among expectant fathers than in a similar group of males whose wives are not pregnant.
2. The offenses are more frequently regressive forms of sexual activity.
3. Sex offenders in the pregnancy group are more likely to be first arrests than nonsex offenders in the same group.
4. Sexual offenses by expectant fathers are more likely to occur in the last 4 months of pregnancy as sexual deprivation increases.
5. Sex offenses are more likely to occur during the wife's first rather than later pregnancies.

METHOD

Subjects

Data are presented on a group of 91 married men examined in a court psychiatric clinic whose arrests occurred during the period of their wives' pregnancy. A sample of 91 married men whose wives were not pregnant, drawn from the same population, was selected as a control group.[1] The two groups were equated as to age, race, and year of offense.

The subjects in the experimental group ranged in age from 18 to 47 with a mean age of 28.9 years. The ages of subjects in the control group ranged from 18 to 45 with a mean age of 29.0 years. There were 22 Negroes and 69 white subjects in each group.

Procedure

The subjects selected for this investigation represented a sampling of the total cases referred to the Psychiatric Institute by the judges of the Municipal Court.

Typically, each patient was examined by a clinical psychologist and a psychiatrist, and relatives or complainants were interviewed by psychiatric social workers.

Initially, data were gathered on all married subjects whose wives were pregnant. Subsequently, a control group was selected from the records of other subjects examined during the same period of time. Information collected included social history, behavioral and demographic data, and various qualitative observations. The present study presents results extracted from this data.

TABLE 1

NUMBER OF SEX OFFENSES AS RELATED TO WIFE'S PREGNANCY

Group	Sex offenses	Nonsex offenses	χ^2	*df*	*p*
Expectant fathers	41	50	15.96	1	.01
Nonexpectant married men	16	75			

Results

The number of sex offenses in the experimental group was far greater than in the control group. Forty-one of the expectant fathers (45%) were arrested for sex offenses as opposed to 16 subjects (17.6%) of the control group. The difference in number of sex offenses is statistically significant ($\chi^2 = 15.96$, $df = 1$, $p < .01$).

The data give evidence that the wife's pregnancy is a stress condition upon the husband such that he responds in the form of antisocial behavior, frequently of a sexual nature. However, the specific type of sexual offense could not be compared because of the small number of sex offenders among the control group. Inspection of the type and frequency of offenses in the expectant-father group does suggest regressive kinds of sexual behavior. Exhibitionism and pedophilia account for more than two thirds of the sexual offenses.

Of the 41 sexual offenses in the experimental group, 17 involved exhibitionism, 11 pedophilia, 5 rape or attempted rape, 4 lewd phone calls and letters, 1 homosexual behavior, 1 transvestitism, 1 possessing pornography, and 1 voyeurism.

Analysis of the frequencies of first arrests in the two groups, reveals 36 persons in the expectant father groups and 41 in the control group. This difference is not statistically significant ($\chi^2 = .379$, $df = 1$, not significant at .50). Further analysis does indicate that among the total group of married men, sexual offenders are more likely to be first offenders than nonsexual offenders ($\chi^2 = 8.65$, $df = 1$, $p < .01$). The frequency of sexual offenders among first arrests does not differ significantly between the expectant fathers and the control group ($\chi^2 = .53$, $df = 1$, not significant at .30).

---

[1] John B. Martin and Jesse Hurley assisted in gathering data and in the statistical analysis.

One apparent explanation of the increase in sexual offenses by expectant fathers is that of sexual deprivation. No direct test of this hypothesis could be made since it would have required data on types and frequencies of sexual outlets during the pregnancy period. The only relevant information available here was a comparison of the frequency of sex offenses in the early as against later period of pregnancy. No significant difference was found.

Among the 41 expectant fathers involved in sex offenses, 16 had wives in their first pregnancy. This compares with 12 of the 50 expectant fathers not involved in sex offenses. The difference between the two groups reaches statistical significance ($\chi^2 = 3.849$, $df = 1$, $p < .05$).

Discussion

The greater frequency of sex offenses among the expectant fathers is interpreted as supporting the hypothesis that pregnancy is a stress upon the father. It appears plausible that persons with weak impulse controls develop strong anxiety which might lead to sexual reactions such as exhibitionism and pedophilia. This stress may have diverse sources. Obviously, pregnancy represents increased responsibility and potential demands economically as well as emotionally. Males who are inadequate may anticipate the child as an intolerable burden and as a pressure upon them to demonstrate greater adequacy.

Pregnancy also has implications for a change in the family constellation. Some men may react to the prospective child in terms of earlier family patterns of competition or jealousy in relation to parents or siblings. Pregnancy emphasizes the maternal role of the wife. This may reinforce the husband's associations or identification with a mother figure as against a wife image, with consequent disturbance of sex reactions, or a reinstatement of conflicts and inhibitions centering around the mother. Sexually deviant behavior in these terms may be viewed as regressive immature adjustive reactions to anxiety. These are compulsive forms of sexual deviation which appear to reduce anxiety about masculinity rather than provide sources of nurturance or normal sexual expression.

The data also indicate a greater frequency of sex offenses in the experimental group during the wife's first pregnancy. This raises several questions. Is there a changed adaptation to subsequent pregnancies? Does the initial arrest represent an isolated incident or do these persons simply find other channels for sexual impulses?

REFERENCES

CURTIS, J. L. A psychiatric study of 55 expectant fathers. *United States Armed Forces Medical Journal*, 1955, **6**, 937–950.

FREEMAN, T. Pregnancy as a precipitant of mental illness in men. *British Journal of Medical Psychology*, 1951, **24**, 49–54.

TOWNE, R. D., & AFTERMAN, J. Psychosis in males related to parenthood. *Bulletin of the Menninger Clinic*, 1955, **19**, 19–26.

WEINER, A., & STEINHILBER, R. The post-partum psychoses. *Journal of the International College of Surgeons*, 1961, **36**, 490–499.

WHITE, M. A., PROUT, C. T., FIXSEN, C., & FONDEUR, M. Obstetricians role in post-partum mental illness. *Journal of the American Medical Association*, 1957, **165**, 138–143.

ZILBOORG, G. Depressive reactions related to parenthood. *American Journal of Psychiatry*, 1931, **10**, 927–962.

(Received August 4, 1965)

Volume 71 August 1966 Number 4

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Chicago

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute

SEYMOUR SARASON
Yale University

JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR Managing Editor	MARY C. MCGARRAGHY Editorial Assistant	HENRY S. SALMON Advertising and Business Manager	ROBERT D. GORMAN Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1966 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 235–246

SOME IMPLICATIONS OF PSYCHOTHERAPY RESEARCH FOR THERAPEUTIC PRACTICE[1]

ALLEN E. BERGIN

Teachers College, Columbia University

A survey of psychotherapy research findings is digested into 6 broad conclusions, and implications for practice and research are drawn from them. They are: (a) Psychotherapy causes clients to become better or worse adjusted than controls, (b) Control Ss improve with time as a result of informal therapeutic encounters, (c) Therapeutic progress varies with therapist warmth, empathy, adjustment, and experience, (d) Client-centered therapy is the only interview-oriented method that has been validated by research, (e) Traditional therapies are seriously limited in effectiveness and are relevant for a small minority of disturbances, and (f) Behavior therapies have considerable promise for enhancing therapeutic effectiveness and should be utilized or experimented with more widely.

The material to follow is a digest of research findings which have implications for practice and research in psychotherapy. It has been formulated in terms of six conclusions and implications which appear justifiable and defensible. This catalogue of conclusions is based upon a comparative handful of research reports which have been carefully selected from the present empirical chaos for their relative adequacy of conceptualization, design, and outcome. Conclusions have been drawn only in those areas where the results appear to have substance and where they have been replicated; consequently, many areas of study are excluded.

THE DETERIORATION EFFECT

Conclusion 1. Psychotherapy may cause people to become better or worse adjusted than comparable people who do not receive such treatment.

Recently, a curious and provocative finding occurred in the preliminary results of the Wisconsin schizophrenia project conducted by Rogers, Gendlin, and Truax (Rogers, 1961; Truax, 1963; Truax & Carkhuff, 1964). It was that the patients in psychotherapy tended to become either better or worse in adjustment than their matched control-group counterparts.

At that time two earlier studies were analyzed (Barron & Leary, 1955; Cartwright, 1956; Cartwright & Vogel, 1960) in which similar findings had occurred; but being incidental to other results, they had not been emphasized in proportion to their true import (Bergin, 1963). Since then, four additional studies with similar findings have been discovered (Fairweather, Simon, Gebhard, Weingarten, Holland, Sanders, Stone, & Reahl, 1960; Mink, 1959; Powers & Witmer, 1951; Rogers & Dymond, 1954). In all seven studies, although there tends to be no difference in the average amount of change between experimentals and controls, there does tend to be a significant difference in *variability* of change. The criterion, or change, scores for treatment groups attain a much wider dispersion than do those of control groups, even though the mean change in both groups is quite similar. Typically, control subjects (*S*s) improve somewhat, with the varying amounts of change clustering about the mean. On the other hand, experimental *S*s are typically dispersed all the way from marked improvement to marked deterioration. Now frequently documented, this information is alarming to say the least. Psychotherapy can and does make people worse than their control counterparts! Because of the controversial nature of this conclusion, the following material is presented

---

[1] Based in part on a paper read at the Pre-Convention Institute of the Ontario Psychological Association, London, Ontario, February 1964. Read at a symposium: "Implications of empirical research for innovations in therapeutic practice and research," American Psychological Association Convention, Los Angeles, September 1964.

TABLE 1

VARIANCES OF DISCREPANCY SCORES ON MMPI SCALES FOR INDIVIDUAL PSYCHOTHERAPY AND NONTREATMENT GROUPS

Scale	Individual psychotherapy ($N = 42$)	Nontreatment group ($N = 23$)	F
	V^a	V	
L	19.89	23.43	1.18
F	215.21	22.94	9.38**
K	55.95	31.70	1.76
Hs	127.46	64.16	1.99*
D	244.30	93.32	2.62**
Hy	113.21	87.80	1.29
Pd	155.00	89.68	1.73
Pa	111.94	68.06	1.64
Pt	208.51	73.27	2.85**
Sc	272.91	74.13	3.68**
Ma	126.79	75.34	1.68
Es	43.56	14.82	2.94**

[a] Variances computed from *SD* data reported by Barron and Leary (Table 2, p. 243).
* $p < .05$.
** $p < .01$.

as detailed substantiating evidence in its support.

Table 1 is reproduced from Cartwright's (1956) reanalysis of the well-known Barron and Leary study (1955).

Cartwright comments on the data as follows:

> For many scales the variance results suggest that mean differences between the groups are absent because differences of two kinds, opposed in sign, are present. It seems that some therapy patients *deteriorated* to a greater extent than did the waiting-list controls, while some therapy patients *did improve* significantly more than the controls [pp. 403–404].

It should be noted that this occurred only for individual and not for group therapy.

It is a fascinating fact that Cartwright's observation has lain unattended in the literature for years, while implicit in his statement is a clear means of resolving much of the controversy over negative results in therapy-outcome studies. It is even more fascinating that Cartwright himself participated in a study (Rogers & Dymond, 1954) in which a similar phenomenon occurred, but just as with the data in the Barron and Leary study it was never emphasized in proportion to its true import. The classic features in this study apparently overshadowed the passing references to a *client-deterioration phenomenon.* While the study is properly famous for other reasons, it provides supporting bits of evidence for the thesis that negative change in therapy is not an isolated or chance occurrence. A careful reading of the report indicates that of 25 therapy *S*s, 6 or 24%, declined in self-ideal correlation between pretherapy and follow-up testing. A quick computation of the mean change in self-ideal correlation indicates that those who increased averaged an increment of .49 in their correlations, whereas those who declined a decrement of −.40, a difference that is striking considering the fact that the mean pretherapy correlations were not different for these two subgroups. While some chance fluctuations in scores are to be expected, these changes in both directions can hardly be attributed to the effects of imperfect test reliability. While Butler and Haigh (1954) do not examine these possibilities in the data, they do allude to them in passing: "It is of interest, though it does not bear directly upon the hypothesis, that there has also been a marked increase in the degree of variation of correlations (self-ideal) over this period [p. 63]."

It may be argued, of course, that decline in self-ideal correlation can be an indication of improved adjustment, particularly when the correlation is extremely high as in the case of some paranoid *S*s. However, the pretest correlations of all six *S*s who declined in this study were low, ranging from .28 to −.12. The question of whether self-ideal correlations actually measure adjustment at all is still a subject of some debate, so it would seem unwise to draw conclusions about psychotherapy in general from data based on this measure alone. In another section of Rogers and Dymond, an analysis of behavior observations made of the clients independently of therapist progress ratings yielded results similar to those found with the self-ideal measure:

> During the whole period from pre-therapy to follow-up, observers saw a definite increase in the maturity of behavior of those clients whose therapy was rated as successful and a sharp decrease in the maturity of behavior of those clients rated as unsuccessful. The relationship was statistically significant [p. 228].

While there are additional fragmentary evidences of deterioration phenomena in the book, these suffice to illustrate the point.

In a controlled study of counseling with high school students, Mink (1959) observes the same phenomenon: "Counseling affected the expression of social adjustments on the California Test of Personality. The forms of expression indicate both improvement and recession [p. 14]."

The excellent multifactor design executed by Fairweather et al. (1960) yielded similar results:

> Generally, significantly different variances occurred on most instruments between treatments and diagnoses. The control group usually had the smallest variance and the three psychotherapy groups the largest [p. 24]. In these three interactions, one or all of the three long-term psychotic groups in psychotherapy demonstrated changes in the maladaptive scale direction [MMPI] while the controls remain[ed] relatively the same or change[d] in the adaptive direction [p. 9].

Cartwright and Vogel (1960) discovered the same type of differential effect in a neurotic sample using different criterion measures:

> Thus, as measured by the Q score, adjustment changes, regardless of direction, were significantly greater during a therapy period than during a No-Therapy period [p. 122]. The Post-therapy tests showed those in therapy with experienced therapists to have improved significantly on both tests, whereas those in therapy with inexperienced therapists not to have improved . . . , in fact they bordered on a significant decrease in health on the TAT [p. 127].

Turning back several decades to the Cambridge-Somerville youth study (Powers & Witmer, 1951) which was initiated in 1937, the same phenomenon is found with a group of predelinquent boys:

> When the Study Services were effectual most of the boys did function better than their C-twins. This conclusion can be accepted, however, only if its opposite is also accepted: that some of the boys who were not benefited may have been handicapped in social adjustment by the organization's efforts. If this is true, we can conclude that the apparent chance distribution of terminal adjustment ratings . . . was due to the fact that the good effects of the Study were counterbalanced by the poor [p. 455].

Elsewhere the authors indicate that in a significant proportion of cases where the counselor's efforts were judged as poor, the boys "were more socially maladjusted than their control twin [p. 509]." It is unfortunate that this excellently designed and executed study is one leaned upon most heavily by Eysenck (1960, 1965) in his bold denial of the usefulness of psychotherapy, for while the study shows no difference between experimentals and controls, it demonstrates the efficacy of treatment as well as its deteriorative effect.

Finally, to cite the recent Wisconsin project on therapy of schizophrenia which has been published (Truax, 1963) thus far only in tempting bits and pieces:

> High levels of therapist-offered conditions during therapy are related to patient improvement, but . . . low levels . . . are related to patient deterioration, so that if all the therapy combined is indiscriminately compared to control conditions there is little average change. Thus, psychotherapy can be for better or for worse [p. 256].

Since the length of therapy varied in these seven studies from a few months to several years, it seems doubtful that the observed deterioration can be accounted for by the temporary regression that sometimes occurs during treatment. The views of most writers would indicate that the average deterioration due to this effect for a treatment group would be small after brief and lengthy periods of therapy but large in between; whereas the findings reported here suggest a consistent, rectangularly distributed, amount of regression, regardless of the length of time transpired prior to obtaining outcome estimates. Unfortunately, so little controlled empirical work has been done with analytic therapies, which are presumably the richest sources of such data, that it is difficult to compare the findings reported here with what might be found if research were done on them.

Fortunately, these various data indicate that psychotherapy can make people considerably better off than control *S*s. Therefore, contrary to the notions of some critics, psychotherapy can produce improvement beyond that which may occur due to spontaneous remission alone. Consistently replicated, this is a direct and unambiguous refutation of the oft-cited Eysenckian position (Eysenck, 1960, 1965).

A general paradigm is suggested by the double-edged effect observed in the studies cited which may be schematized as shown in Figure 1. Such a startling phenomenon certainly deserves a name, and *The Deterioration Effect* is suggested here.

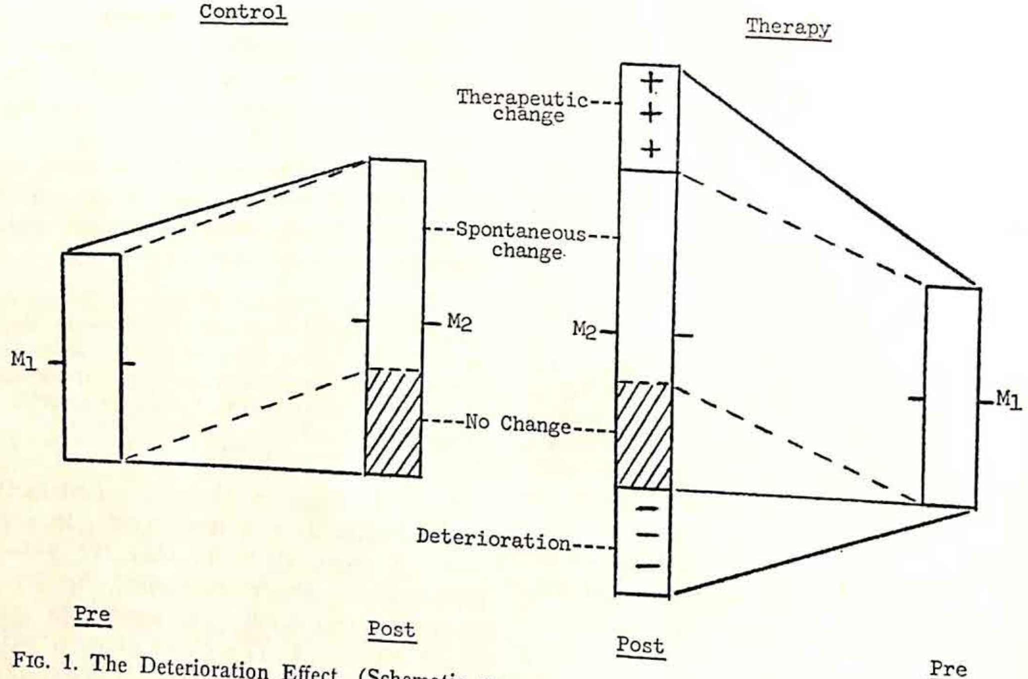


FIG. 1. The Deterioration Effect. (Schematic representation of pre- and posttest distributions of criterion scores in psychotherapy-outcome studies.)

It is interesting to note that a phenomenon similar to the great variability in the quality of therapeutic effects noted here has also been observed in relation to the accuracy of diagnostic evaluations (Garfield, 1963). Apparently, even well-known diagnosticians vary greatly in the accuracy of their judgments. When all of these judgments are pooled, average predictions or discriminations often are not different from chance estimates; but some individuals appear to far exceed chance predictions while others actually do worse than chance.

Implication 1. (*a*) The practice of psychotherapy should not be given up as some have advocated. (*b*) Those engaged in this field should be more cautious and critical of their own practices, carefully eliminationg any ineffective or harmful therapeutic techniques. They should find out whom they are making worse or better, and how, with all due speed. (*c*) They should find out if some therapists make people better and if some make them worse, or if individual therapists do both. After that, comes the ticklish business of making changes in technique, personality, or personnel as may be necessary to eliminate negative influences and accentuate positive ones.

NATURAL THERAPEUTIC CONDITIONS

Conclusion 2. (*a*) It has been frequently replicated, and is now a well-established fact, that control *Ss* who do not receive psychotherapy change positively as a group with the passage of time. This is the so-called "spontaneous remission" effect (Eysenck, 1952, 1960, 1965). (*b*) Three studies (Frank, 1961; Gurin, Veroff, & Feld, 1960; Powers & Witmer, 1951) indicate that many of these disturbed persons who receive no formal psychotherapy seek and obtain help from various professional and nonprofessional sources such as friends, clergymen, physicians, teachers, and occasionally even psychotherapists (Bergin, 1963).

All this has typically been unknown to the researchers who were depending upon these so-called controls to be a baseline for comparison with their treatment cases. It seems clear that this aid has an ameliorative effect, as the people improve, although it would be impossible to substantiate this fully without further study of the influences upon control *Ss* in their "natural" habitat. To the extent that this position is correct, it further under-mines the Eysenck-type position, because it shows that control *S*s often change due to the influence of therapy or therapy-like procedures. Thus, "spontaneous remission" is just another name for the effects of informal therapy.

Implication 2. (*a*) Researchers who utilize control groups should carefully ascertain that these groups are indeed controls, or, if necessary, should directly measure the effects of nonexperimental influences which they cannot control. (*b*) The fact that some of these previously uncontrolled influences are much like therapy, but frequently occur outside of a professional setting, implies that nonprofessional help can stimulate positive personality change. This may consist partly of individuals with "therapeutic personalities" who are sought out for counsel and catharsis by many people. It may be also that unrecognized, but powerful, therapeutic agents exist naturally in everyday life. Just as cures for various physical disorders have been discovered by studying health, so it may be possible to discover antidotes for some of the mental disorders that confront us by discovering conditions already existing in "nature" which support or promote personality integration.

INGREDIENTS OF THERAPY

Conclusion 3. Therapeutic progress varies as a function of therapist characteristics such as warmth, empathy, adequacy of adjustment, and experience.

In a recent review, Gardner (1964) cited a smattering of positive results to the effect that the more a therapist has an attitude of *liking and warmth* the more likely he is to obtain positive change in his clients. While some of the studies enumerated are of questionable design or generalizability, they are relatively consistent when compared with many other areas of research.

A recent questionnaire study of patients' retrospective reports regarding their therapeutic experience (Strupp, Wallach, & Wogan, 1964), which was not reported by Gardner, further confirms this general finding. While the study is uncontrolled and appears to be contaminated by artifactually inflated correlations, it is of interest that it strongly emphasizes the importance of therapist warmth and genuineness in relation to patient-perceived outcome ($r = .53$).

Additional data on this point come from the client-centered group in a series of studies with neurotics and psychotics. It should be noted that some of the therapists studied were *not* client-centered. These studies are consistent in discovering a significant relationship between operational measures of Rogers' concept of positive regard and independent indices of therapeutic progress or outcome (Truax & Carkhuff, 1964, 1965a; Barrett-Lennard, 1962). Measures of the therapist's attitudes have included ratings by both the therapist himself and the patient. Three types of analysis have resulted in similar findings and in different studies with different samples of clients and therapists. It has thus become increasingly clear, within the limits of these studies, that a therapist's ability to be warm and positively inclined toward his patients is an effective therapeutic ingredient. The effects of intentional authoritarian demands or other forms of planned therapist aggression which are sometimes advocated have not been studied and thus cannot be compared with these findings.

Acknowledging the past confusion and contradiction involved in studies of *empathy,* it is suggested that the recent data summarized at Chicago (Barrett-Lennard, 1962), Wisconsin (Truax, 1961b; Truax & Carkhuff, 1964), and Kentucky (Dickenson & Truax, 1965; Truax & Carkhuff, 1965a; Truax, Carkhuff, & Kodman, 1965; Truax & Wargo, 1965) offer promising leads. Analyses of recorded therapist behavior and ratings by clients of their therapists during the process of treatment have yielded consistently positive relationships between empathic understanding and outcome.

The strength of these findings lies in careful design (Rogers, 1961) and in the analysis of therapist behavior *in vivo,* which is unusual in empathy research. A new empathy measure has been operationalized by Truax (1961b) and is defined by Truax and Carkhuff (1964) as accurate "sensitivity to current feelings *and* the verbal facility to communicate this understanding in a language attuned to the patient's current being [p. 8]." While the scale is still crude and might not be accepted by analysts as measuring their "kind" of empathy, its usefulness has been relatively substantial in these studies.

The third characteristic, *adequacy of adjustment,* has not been studied as thoroughly as the others, but thus far the data are relatively consistent. Those therapists who are more anxious, conflicted, defensive, or "unhealthy" are least likely to promote change in their cases.

Several studies have indicated that supervisor and client ratings of the therapists' competence are negatively related to his degree of anxiety or maladjustment (Arbuckle, 1956; Bandura, 1956; Bergin & Solomon, 1963). Other studies have yielded similar findings when the therapist's actual in-therapy behavior and the patient's response to it was evaluated and used as a criterion of competence. For example, Bandura, Lipsher, and Miller (1960) found that therapists' hostility anxiety was directly associated with avoidance responses to patients' expressions of hostility toward them. The more hostility conflict a therapist had, the more likely he was to avoid his patients' hostility and consequently the patient's self-exploration in this area diminished and his conflicts remained unresolved. A practically identical result was found by Winder, Ahmad, Bandura, and Rau (1962) with regard to dependency anxiety.

In another study (Bergin & Solomon, 1963) it was found that measures of the therapists' degree of personal disturbance correlate negatively with his level of empathy as measured by ratings of tape-recorded psychotherapy interviews. Independent measures of personality strength, on the other hand, correlated positively with degree of "live" empathy. In addition, ratings of therapist anxiety level correlated negatively with independent ratings of therapeutic competence.

Additional data come from the client-centered studies already cited with regard to warmth and empathy, in their examination of therapist congruence. Congruence (Rogers, 1957, 1959) means essentially the healthiness of the therapist in his relationship with his client—his spontaneity, nondefensiveness, openness, or genuineness. Like positive regard and empathy, this variable has also been related to therapeutic progress, and further confirms the general finding of a direct connection between level of therapist adjustment and therapeutic effectiveness.

The three elements of warmth, empathy, and congruence have been found, in the Wisconsin studies, to vary directly with outcome in both negative and positive directions. That is, when these therapist characteristics were at a low level, the patients were getting worse; when they were high, the patients improved (Truax & Carkhuff, 1964). These studies thus provide a partial answer to the question raised earlier as to how negative change occurred in the outcome studies reviewed, although they are limited in that the observed differences were not large, and there is also some question as to whether the division into high and low conditions was done before or after the fact. The other studies cited here in the same realm further clarify the point, although none of the data are precise enough to make practical selection decisions possible.

With regard to the much debated variable of therapist experience, it may be asserted that, in general, more experienced therapists are more effective and successful. This is based on four studies (Barrett-Lennard, 1962; Cartwright & Vogel, 1960; Chance, 1959; Fiedler, 1950a, 1950b, 1951), one of which suggests that highly inexperienced therapists may actually cause patient deterioration (Cartwright & Vogel, 1960).

Implication 3. (*a*) Since psychotherapists are effective partly as a function of personal adjustment, they should be selected for this quality and not solely on the basis of academic and intellectual abilities. Future practice of therapy should therefore be modified by new selection procedures which will bring healthier personalities to bear upon problems of pathology, and by closer self-scrutiny and exposure of one's work among present practitioners.

There is presently no evidence that personal therapy for a disturbed therapist can qualify him for practice and should not be depended upon to perform that function until such evidence is provided. This does not, of course, prove that the experience of being treated cannot be useful to a student therapist whose functioning is within a relatively normal range. There are no studies in which treated neurotics have improved to a level of functioning which is similar to that of control normals even though they do change in level of adjustment; therefore, treatment should not be counted upon to take care of errors in selection. The behavior ratings and personality inventories used in the studies reviewed could provide a beginning in research geared specifically toward the selection problem.

(*b*) Given the necessary personal attributes, therapists should develop their abilities in the realm of warmth and empathic communication, particularly in the case of empathy which is known to be subject to training and experience influences. Further study should be conducted so that clear, measurable standards of performance can be required of aspirants to professional status before they are permitted to practice. As an example, the Truax Empathy Scale (Truax, 1961b) could be used as a beginning to assess one's level of functioning via analysis of recorded interviews.

(*c*) Inexperienced potential therapists should be very carefully introduced to practice with clients, perhaps with much more stringent care than is now commonly exercised. Since all beginners make many mistakes, it may be useful and ethical to have them see more resilient, normal people until they reach a criterion level of interview performance, measured perhaps on dimensions such as warmth and empathy which appear to be accepted by most schools of therapy as vital though not necessarily sufficient for successful treatment.

Conclusion 4. To date, the only school of interview-oriented psychotherapy which has consistently yielded positive outcomes in research studies is the client-centered approach (Rogers & Dymond, 1954; Shlien, Mosak, & Dreikurs, 1962; Truax & Carkhuff, 1964).

The fact that other schools have not subjected their methods to systematic study of this sort is important but it should not deter one from accepting the fact that client-centered treatment has some positive value when properly conducted according to Rogers' (1957) paradigm. The implications for practice seem quite clear, particularly in view of the consistently dismal reports on percentages of improvement in psychoanalytic therapy (Eysenck, 1965; Wolpe, 1964b).

It appears from these reports that the poorest results were obtained with more classical, long-term psychoanalysis, namely a lower percentage of improved cases than the 67% "spontaneous" remission rate. Briefly, analytically oriented eclectic psychotherapy was more promising in that the percentage improvement equaled the spontaneous remission figure. This type of therapy was also used in some of the studies cited in this paper on the deterioration effect; therefore, despite the generally negative evidence, some analytically oriented therapists must be having a positive effect beyond that occurring in control groups.

It should also be noted that the technique of "moderate interpretation" (Speisman, 1959), which derives from the analytic tradition, has potential therapeutic significance. Its definition is very similar to that given for "good" interpretation by various analysts (Fenichel, 1941) and it is related to productive patient self-exploration. It consists of responding to client affect just below the surface and labeling, identifying, or emphasizing it. This does not involve making connections between past and present, being diagnostic or theoretical, nor telling the patient about feelings he "really has" when he's not experiencing them. It is, rather, an instance of good empathy. If one looks carefully at the definitions and operations for identifying accurate empathy and moderate or good interpretation, it is very difficult to distinguish between them. Truax and Carkhuff (1964) refer to this notion in an interesting comment:

> "accurate empathy" has much in common with the "good psychoanalytic interpretation," in that it makes use of both verbal and nonverbal cues presented by the patient. It differs from some good psychoanalytic interpretations in its insistence that the therapist's empathic response focuses upon feelings and experience of the patient from the patient's own unique viewpoint.

The importance of these observations should not be underestimated, for if they are accurate it appears that effective variables cut across schools of treatment and thus provide the basis for applying techniques on the basis of known effects rather than on doctrines promulgated by warring factions. This also indicates that titles, degrees, or years of training should not define the psychotherapist, but rather what the individual can do. Thus one might call himself "client-centered" and espouse the teachings of that school while at the same time presenting the low level of therapist empathy found to result in client deterioration. On the other hand, a psychoanalyst might be functioning at a high level according to the client-centered empathy scale.

Conclusion 5. In spite of all so far stated about the possibilities for substantially improving consulting-room effectiveness, some stubborn facts still require confrontation. One is that even when the various sources of slippage and inadequacy are accounted for, interviews still do not generally produce very dramatic changes in people. Another is the now well-known fact that many types of people simply are not helped at all by this procedure.

Studies of the relationship between client qualities and therapeutic outcome indicate consistently and clearly that positive outcome is limited or nil with many personality types. It is common for private practitioners and even clinics either to refuse to treat, or reluctantly to accept for treatment, cases that do not fit their conception of psychotherapy. To a great extent this is realistic because traditional methods do not work with these cases. These "rejects," as compared with "accepted" cases, tend to be less intelligent, less anxious, less educated, less verbal and insightful, more concrete and action-oriented, more severely disturbed, more impulsive in the sociopathic sense, and often find the typical consulting-room procedure rather meaningless (Barron, 1953; Cartwright, 1955; Fulkerson & Barry, 1961; Garfield & Affleck, 1961; Hollingshead & Redlich, 1958; Kirtner & Cartwright, 1958a, 1958b). This general observation has been made fairly frequently by various clinicians and is currently rather well-substantiated by the research literature.

Implication 5. The implication of these data, which only confirm an already widely believed idea, is that novel or modified techniques must be developed for dealing with a vast population whose problems are not amenable to standard methods. The importance of novel approaches is further emphasized by the fact that standard methods are not dramatically effective even in those cases where they are applicable, except in rare instances. The latter unusual cases would be a proper subject of study in themselves and may actually suggest innovations even though they arise in "traditional" therapy.

There are three primary sources of possible innovation that might alleviate this predicament. One is creative work in the clinical setting; another is naturally existing conditions in society; and another is that general area of research which is concerned with personality and behavior change such as studies of learning, attitude change, and personality development.

THE PROMISE OF BEHAVIOR THERAPY

Conclusion 6. Studies of learning have thus far been very fruitful in generating principles and methods for promoting personality change. The work by Wolpe (1958), Lazarus (1963), Lang and Lazovik (1963), Lindsley (1963), and others has been both provocative and fruitful. The cases presented and research studies reported provide more positive evidence of the usefulness of these methods than is the case in any form of traditional interview or dynamic psychotherapy, including client-centered therapy.

They involve clinical adaptation of learning principles, such as counterconditioning or extinction of anxiety symptoms, positive reinforcement in shaping adaptive responses and developing appropriate discriminations, aversive conditioning of maladaptive approach responses, and modeling. It is the effects of these methods which are important here. Wolpe (1964a) cites over 200 cases of neurosis in 89% of which he has obtained substantial recovery. Lazarus (1963), in England, reports 408 cases with a similar improvement rate. The striking aspect of these results is that they have been achieved with difficult symptom pictures in brief periods of time. Unfortunately, these are clinical reports by individual therapists who rate their own case outcomes. Independent criteria and control *S*s are completely lacking, and it is difficult to discern how comparable their cases are with those reported in other studies. Still, it is rare to find such high rates of claimed cure even in the clinical literature.

A number of well-designed studies appear to substantiate the clinical reports of Wolpe and Lazarus. Lang and Lazovik (1963) were able significantly to alter snake phobias with brief desensitization procedures. Effects of testing and training in relaxation were controlled, and no symptom substitution occurred during 6 months of follow-up. Lazarus (1961) demonstrated substantial and rapid change of phobic symptoms and impotence by group desensitization methods. A comparison group being treated by traditional interpretive group therapy showed considerably less improvement, only 2 of 17 cases becoming symptom free after 22 sessions. These same cases were subsequently treated by group desensitization and after a mean of 10 sessions each, two thirds were symptom free. Paul (1966) found that desensitization procedures were far more effective in eliminating speech anxieties than brief insight therapy, an attention-placebo condition, and a no-therapy control condition.

In a study of operant conditioning methods, which are different from Wolpe's techniques, King, Armitage, and Tilton (1960) found that substantial changes could be effected even in schizophrenic cases. They were able to produce clinically observable improvement in cases so treated which was greater than the changes occurring in conventional interview therapy, recreational therapy, or no therapy. Ayllon and Michael (1959) effected substantial positive changes in ward behavior of psychotics by programing the reinforcements of their hospital environment according to operant principles. Lovaas, Schaeffer, and Simmons (1966) appear to have induced important changes in the social behavior of difficult cases of childhood autism by systematic use of negative reinforcement. In a review, Lindsley (1963) argues for the general promise of operant techniques; although the evidence thus far pertains primarily to simple motor and verbal behaviors. Conceivably, this approach will prove to be more useful with the more primitive behaviors of psychotics and small children than with the more complex, symbolically involved adult neuroses.

A most interesting development in behavior therapy involves the systematic application of principles of imitative or observational learning. Bandura (1965b) argues persuasively from the vantage point of extensive experimental work (Bandura, 1965a, 1965c) that modeling procedures provide powerful conditions for the acquisition of new responses and the modification of old ones. Though controlled clinical applications have just begun, they already lend considerable substance to Bandura's view (Berberich & Schaeffer, 1965; Frank, 1965; Hoehn-Saric, Frank, Imber, Nash, Stone, & Battle, 1965; Krumboltz & Thoreson, 1964; Krumboltz & Schroeder, 1965; Krumboltz, Varenhorst, & Thoreson, 1965; Nelson & Bijan, 1965; Thoreson & Krumboltz, 1965; Truax & Carkhuff, 1965b).

Several extensive reviews further substantiate the generality of Conclusion 6 (Bandura, 1965b; Bandura & Walters, 1963; Eysenck & Rachman, 1965; Franks, 1964; Grossberg, 1964; Krasner & Ullman, 1965; Ullman & Krasner, 1965; Wolpe, 1964b).

In spite of the fact that the evidence is favorable, these techniques have been criticized by clinicians as removing symptoms without changing basic pathology and as being limited to very simple neuroses. Neither criticism, however, fits the evidence. Wolpe (1964a) cites data on 88 cases which indicate that a high proportion of complex neuroses can be successfully treated (89%) and in a much briefer time than is typical of traditional methods (Table 2).

TABLE 2

COMPARISON OF NUMBERS OF SESSIONS IN COMPLEX AND SIMPLE NEUROSES

Neuroses	*N*	Mdn No. sessions	*M* No. sessions
Complex	65	29	54.8
Simple	21	11.5	14.9
Total	86	23	45.4

Note.—The total is only 86 because 2 cases that turned out to be schizophrenic are excluded.

The more telling critique of this work is Breger and McGaugh's (1965) point regarding the uncontrolled case reports, which are the basis for the high cure rates, and the rater bias in estimating outcomes encountered in many of the experimental studies. Faulty as a proportion of these reports are, the overall record still represents the best there is in the field of psychotherapy.

In addition to the fact that difficult cases show improvement in a short time, these reports indicate that significant relapses are rare. This is perhaps the most persuasive evidence that behavior therapists are right when they assert that "symptoms" are not symptoms of psychoanalytic-style pathology, but that they are learned behaviors subject to modification via relearning.

Some learning theorists have criticized Wolpe in particular, claiming that his techniques do not derive directly and logically from learning principles and thus do not have the scientific base he claims (Breger & McGaugh, 1965; Mowrer, 1963). While this may be true to some extent, it is irrelevant to the question of the technique's effectiveness and ignores the possibility that these clinical phenomena may eventually become the basis for reformulating learning theories in terms of complex, socially significant human behavior. In this case, one would not expect principles of behavior therapy to conform rigorously to conceptions derived largely from animal research.

Implication 6. The implications of this work seem quite clear. Since these techniques are effective with many types of symptomatology, they should be used. With regard to some of the more complex and difficult problems, behavior therapists argue that it would be better to spend time developing more complex social learning paradigms for treatment than to expend equal energy modifying less promising traditional interview methods. It appears that special effort should be devoted to integrating these methods with others and in some cases substituting them for the other methods. It would seem important to avoid a current tendency to isolate behavior therapies from the mainstream of treatment and thus create another rigid "school" which will gradually become as impervious to new ideas as the traditional schools already are.

CONCLUSION

In conclusion, it is only regrettable that comment upon so many topics of research has had to be excluded. Suffice it to say that the results in many of those not mentioned are not as yet amenable to synthesis. A good example is the material on the patient-therapist relationship. Nearly all of this research actually pertains to therapist qualities and has nothing to do with an analysis of interactional factors. An unusual exception is the work of Barrett-Lennard (1962) which was cited briefly in the discussion of therapist qualities. The few other useful facts in this domain were also included in that section. Another promising line of investigation is that on patient-therapist similarity; but the meaning of the data is still quite ambiguous (Sussman, 1964).

In spite of the fact that much of what is called psychotherapy research is appalling in its inadequacy, to have found a handful of reliable conclusions is gratifying. The groundwork seems well laid by these studies for initial steps at productive innovation in therapeutic treatment.

REFERENCES

ARBUCKLE, D. S. Client perception of counselor personality. *Journal of Counseling Psychology,* 1956, **3**, 93–96.

AYLLON, T., & MICHAEL, J. The psychiatric nurse as a behavioral engineer. *Journal of the Experimental Analysis of Behavior,* 1959, **2**, 323–334.

BANDURA, A. Psychotherapist's anxiety level, self-insight, and psychotherapeutic competence. *Journal of Abnormal and Social Psychology,* 1956, **52**, 333–337.

BANDURA, A. Behavioral modification through modeling procedures. In L. Krasner & L. Ullmann (Eds.), *Research in behavior modification.* New York: Holt, Rinehart & Winston, 1965. Pp. 310–340. (a)

BANDURA, A. Psychotherapy conceptualized as a social-learning process. Paper read at the Kentucky Centennial Symposium on Psychotherapy, University of Kentucky, April 1965. (b)

BANDURA, A. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965. Pp. 3–48. (c)

BANDURA, A., LIPSHER, D. H., & MILLER, P. E. Psychotherapists' approach-avoidance reactions to patients' expressions of hostility. *Journal of Consulting Psychology,* 1960, **24**, 1–8.

BANDURA, A., & WALTERS, R. H. *Social learning and personality development.* New York: Holt, Rinehart & Winston, 1963.

BARRETT-LENNARD, G. T. Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs,* 1962, **76**(43, Whole No. 562).

BARRON, F. Some test correlates of response to psychotherapy. *Journal of Consulting Psychology,* 1953, **17**, 235–241.

BARRON, F., & LEARY, T. Changes in psychoneurotic patients with and without psychotherapy. *Journal of Consulting Psychology,* 1955, **19**, 239–245.

BERBERICH, J., & SCHAEFFER, B. Establishment of verbal behavior through imitation. Paper read at

American Psychological Association, Chicago, September 1965.

Bergin, A. E. The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology*, 1963, **10**, 244–250.

Bergin, A. E., & Solomon, S. Personality and performance correlates of empathic understanding in psychotherapy. *American Psychologist*, 1963, **18**, 393. (Abstract)

Breger, L., & McGaugh, J. L. Critique and reformulation of "learning-theory" approaches to psychotherapy and neurosis. *Psychological Bulletin*, 1965, **63**, 338–358.

Butler, J. M., & Haigh, G. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1954. Pp. 55–75.

Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *Journal of Consulting Psychology*, 1955, **19**, 357–363.

Cartwright, D. S. Note on "changes" in psychoneurotic patients with and without psychotherapy. *Journal of Consulting Psychology*, 1956, **20**, 403–404.

Cartwright, R. D., & Vogel, J. L. A comparison of changes in psychoneurotic patients during matched periods of therapy and no-therapy. *Journal of Consulting Psychology*, 1960, **24**, 121–127.

Chance, E. *Families in treatment*. New York: Basic Books, 1959.

Dickenson, W. A., & Truax, C. B. Group counseling with college underachievers: Comparisons with a control group and relationship to empathy, warmth, and genuineness. Unpublished manuscript, University of Kentucky, 1965.

Eysenck, H. J. The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 1952, **16**, 319–324.

Eysenck, H. J. The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1960. Pp. 697–725.

Eysenck, H. J. The effects of psychotherapy. *International Journal of Psychiatry*, 1965, **1**, 97–178.

Eysenck, H. J., & Rachman, S. *The causes and cures of neurosis*. San Diego: Knapp, 1965.

Fairweather, G. W., Simon, R., Gebhard, M. E., Weingarten, E., Holland, J. L., Sanders, R., Stone, G. B., & Reahl, J. E. Relative effectiveness of psychotherapeutic programs: A multicriteria comparison of four programs for three different patient groups. *Psychological Monographs*, 1960, **74**(5, Whole No. 492).

Fenichel, O. *Problems of psychoanalytic techniques*. Albany: Psychoanalytic Quarterly, 1941.

Fiedler, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective, and Adlerian therapy. *Journal of Consulting Psychology*, 1950, **14**, 436–445. (a)

Fiedler, F. E. The concept of the ideal therapeutic relationship. *Journal of Consulting Psychology*, 1950, **14**, 239–245. (b)

Fiedler, F. E. Factor analyses of psychoanalytic, nondirective, and Adlerian therapeutic relationships. *Journal of Consulting Psychology*, 1951, **15**, 32–38.

Frank, J. D. *Persuasion and healing*. Baltimore: Johns Hopkins Press, 1961.

Frank, J. D. The role of hope in psychotherapy. Paper read at the University of Kentucky Centennial Psychotherapy Symposium, April 1965.

Franks, C. (Ed.) *Conditioning techniques in clinical practice and research*. New York: Springer, 1964.

Fulkerson, S. D., & Barry, J. R. Methodology and research on the prognostic use of psychological tests. *Psychological Bulletin*, 1961, **58**, 177–204.

Gardner, G. G. The psychotherapeutic relationship. *Psychological Bulletin*, 1964, **61**, 426–437.

Garfield, S. L. The clinical method in personality assessment. In J. Wepman & R. Heine (Eds.), *Concepts of personality*. Chicago: Aldine, 1963. Pp. 474–502.

Garfield, S. L., & Affleck, D. C. Therapists' judgments concerning patients considered for psychotherapy. *Journal of Consulting Psychology*, 1961, **25**, 505–509.

Grossberg, J. M. Behavior therapy: A review. *Psychological Bulletin*, 1964, **62**, 73–88.

Gurin, G., Veroff, J., & Feld, S. *Americans view their mental health*. New York: Basic Books, 1960.

Hoehn-Saric, R., Frank, J. D., Imber, S. D., Nash, E. H., Stone, A. R., & Battle, C. C. Systematic preparation of patients for psychotherapy: I. Effects on therapy behavior and outcome. *Journal of Psychiatric Research*, 1965, **2**, 267–281.

Hollingshead, A. B., & Redlich, F. C. *Social class and mental illness*. New York: Wiley, 1958.

King, G. F., Armitage, S. G., & Tilton, J. R. A therapeutic approach to schizophrenics of extreme pathology. *Journal of Abnormal and Social Psychology*, 1960, **61**, 276–286.

Kirtner, W. L., & Cartwright, D. S. Success and failure in client-centered therapy as a function of client personality variables. *Journal of Consulting Psychology*, 1958, **22**, 259–264. (a)

Kirtner, W. L., & Cartwright, D. S. Success and failure in client-centered therapy as a function of initial in-therapy behavior. *Journal of Consulting Psychology*, 1958, **22**, 329–333. (b)

Krasner, L., & Ullman, L. (Eds.) *Research in behavior modification: New developments and implications*. New York: Holt, Rinehart & Winston, 1965.

Krumboltz, J. D., & Schroeder, W. W. The effect of reinforcement counseling and model-reinforcement counseling on information-seeking behavior of high school students. *Personnel and Guidance Journal*, 1966, in press.

Krumboltz, J. D., & Thoreson, C. E. The effect of behavioral counseling in group and individual settings on information-seeking behavior. *Journal of Counseling Psychology*, 1964, **9**, 324–333.

Krumboltz, J. D., Varenhorst, B., & Thoreson, C. E. Non-verbal factors in the effectiveness of models in counseling. Paper read at American Personnel and Guidance Association, Minneapolis, April 1965.

LANG, P. J., & LAZOVIK, A. D. Experimental desensitization of a phobia. *Journal of Abnormal and Social Psychology,* 1963, **6**, 519–525.

LAZARUS, A. A. Group therapy of phobic disorders by systematic desensitization. *Journal of Abnormal and Social Psychology,* 1961, **63**, 504–510.

LAZARUS, A. A. An evaluation of behavior therapy. *Behavior Research and Therapy,* 1963, **1**, 69–79.

LINDSLEY, O. R. Free-operant conditioning and psychotherapy. In J. H. Masserman (Ed.), *Current psychiatric therapies.* Vol. 3. New York: Grune & Stratton, 1963. Pp. 47–56.

LOVAAS, O. I., SCHAEFFER, B., & SIMMONS, J. Q. Building social behavior in autistic children by use of electric shock. In J. O. Palmer & M. J. Goldstein (Eds.), *Perspectives in psychopathology: Readings in abnormal psychology.* New York: Oxford University Press, 1966. Pp. 222–236.

MINK, O. G. A comparison of effectiveness of nondirective therapy and clinical counseling in the junior high school. *School Counselor,* 1959, **6**, 12–14.

MOWRER, O. H. Freudianism, behavior therapy, and "self-disclosure." *Behavior Research and Therapy,* 1963, **1**.

NELSON, K., & BIJAN, G. Teaching social behaviors to schizophrenic children through imitation. Paper read at American Psychological Association, Chicago, September 1965.

PAUL, G. L. *Effects of insight, desensitization, and attention placebo treatment of anxiety.* Stanford: Stanford University Press, 1966.

POWERS, E., & WITMER, H. *An experiment in the prevention of delinquency.* New York: Columbia University Press, 1951.

ROGERS, C. R. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology,* 1957, **21**, 95–103.

ROGERS, C. R. A theory of therapy, personality, and interpersonal relationships, as developed in the client-centered framework. In S. Koch (Ed.), *Psychology; a study of a science.* Vol. 3. New York: McGraw-Hill, 1959. Pp. 184–256.

ROGERS, C. R. A theory of psychotherapy with schizophrenics and a proposal for its empirical investigation. In J. G. Dawson & N. P. Dellis (Eds.), *Psychotherapy with schizophrenics.* Baton Rouge: Louisiana State University Press, 1961. Pp. 3–19.

ROGERS, C. R., & DYMOND, R. F. *Psychotherapy and personality change.* Chicago: University of Chicago Press, 1954.

SHLIEN, J. M., MOSAK, H. H., & DREIKURS, R. Effect of time limits: A comparison of two psychotherapies. *Journal of Counseling Psychology,* 1962, **9**, 31–34.

SPEISMAN, J. C. Depth of interpretation and verbal resistance in psychotherapy. *Journal of Consulting Psychology,* 1959, **23**, 93–99.

STRUPP, H. H., WALLACH, M. S., & WOGAN, M. Psychotherapy experience in retrospect: Questionnaire survey of former patients and their therapists. *Psychological Monographs,* 1964, **78**(11, Whole No. 588).

SUSSMAN, A. Patient-therapist similarity as a factor in psychotherapy. Unpublished manuscript, Teachers College, Columbia University, 1964.

THORESON, C. E., & KRUMBOLTZ, J. D. Relationship of counselor reinforcement of selected responses to external behavior. *Journal of Counseling Psychology,* 1966, in press.

TRUAX, C. B. A scale for the measurement of accurate empathy. *Psychiatric Institute Bulletin,* Wisconsin Psychiatric Institute, University of Wisconsin, 1961, **1**, No. 10. (a)

TRUAX, C. B. The process of group psychotherapy: Relationships between hypothesized therapeutic conditions and intrapersonal exploration. *Psychological Monographs,* 1961, **75**(7, Whole No. 511). (b)

TRUAX, C. B. Effective ingredients in psychotherapy. *Journal of Counseling Psychology,* 1963, **10**, 256–263.

TRUAX, C. B., & CARKHUFF, R. R. For better or for worse: The process of psychotherapeutic change. In *Recent advances in behavioral change.* Montreal: McGill University Press, 1964.

TRUAX, C. B., & CARKHUFF, R. R. Experimental manipulation of therapeutic conditions. *Journal of Consulting Psychology,* 1965, **29**, 119–124. (a)

TRUAX, C. B., & CARKHUFF, R. R. Personality change in hospitalized mental patients during group psychotherapy as a function of the use of alternate sessions and vicarious therapy pretraining. *Journal of Clinical Psychology,* 1965, **21**, 225–228. (b)

TRUAX, C. B., & CARKHUFF, R. R., & KODMAN, F. Relationships between therapist-offered conditions and patient change in group psychotherapy. Unpublished manuscript, University of Kentucky, 1965.

TRUAX, C. B., & WARGO, D. G. Human encounters that change behavior: For better or for worse. Unpublished manuscript, University of Kentucky, 1965.

ULLMANN, L., & KRASNER, L. (Eds.) *Case studies in behavior modification.* New York: Holt, Rinehart & Winston, 1965.

WINDER, C. L., AHMAD, F. Z., BANDURA, A., & RAU, L. Dependency of patients, psychotherapists' responses, and aspects of psychotherapy. *Journal of Consulting Psychology,* 1962, **26**, 129–134.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford University Press, 1958.

WOLPE, J. Behavior therapy in complex neurotic states. *British Journal of Psychiatry,* 1964, **110**, 28–34. (a)

WOLPE, J. The comparative clinical status of conditioning therapies and psychoanalysis. In J. Wolpe, A. Salter, & L. J. Reyna (Eds.), *The conditioning therapies.* New York: Holt, Rinehart & Winston, 1964. Pp. 5–20. (b)

(Received June 2, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 247–252

EFFECTS OF CHLORPROMAZINE ON PRIMARY-PROCESS THOUGHT MANIFESTATIONS

THEODORE SARETSKY[1]

Brooklyn College

The object of the present study was to determine the effect of chlorpromazine on primary-process thinking. Ss were 40 institutionalized schizophrenics who were randomly assigned to either a drug or placebo group. Research data were obtained from pre- and postdrug Rorschach protocols, which were scored using Holt's (1959) method. The most outstanding finding was the striking improvement in mean defense effectiveness. Drug patients proved to be more capable of handling primary-process elements; the more so, the more they were rated as clinically improved. The insignificant changes in the sum of Level 1 primary process and mean defense demand suggest that the degree of thought pathology itself is not directly affected by drug treatment.

The present research investigates the psychological nature of the effects of chlorpromazine on chronic schizophrenics. Numerous reports indicate that this drug has appreciable effects on the behavioral status of patients. It reduces motor excitement, agitation, noisiness, and aggression. Treated patients show more interest in self-care and become more reliable in handling privileges, work assignments, and money. Unfortunately, however, these findings have been based almost exclusively upon psychiatric observations and subjective clinical impressions as to the changes in overt symptoms. The effect of the drug on intrinsic psychotic mechanisms, such as the presence of uncontrolled primary-process thinking,[2] has not been studied intensively in psychoanalytic theory. Uncontrolled, primary-process thinking is the factor most constantly associated with extreme ego weakness (Bellak, 1958; Rapaport, 1951; Schafer, 1958). Accordingly, it was used here as the measure of basic change.

Only a minority of the numerous clinical studies on chlorpromazine have employed standardized psychological tests to assess drug effects. Some of the most powerful of the projective methods have played only a marginal role, if any, in these studies. Consequently, we know very little about how chlorpromazine affects such psychological functions as primary- and secondary-process thinking. Because changes in the relative weights of these two types of thinking, as influenced by chlorpromazine, are the major interest here, this study emphasized the Rorschach, scored following Holt's (1959) method as described below.

THEORETICAL CONSIDERATIONS

In 1900 Freud (1938) early distinguished between a "primary process," wherein the instinctual drives manifest themselves without control or delay, and a "secondary process," wherein drives are ordered and controlled by rational thought and voluntary action.

Despite considerable theoretical controversy, there is general agreement (Brenner, 1955; Hartmann, 1950; Holt, 1956; Schafer, 1958) that the primary process is genetically and formally the more primitive, operating with unneutralized, mobile drive energies, and regulated by the pleasure principle. It constitutes striving toward immediate gratification without regard for reality or logic and through the mechanisms of displacement, condensation, substitute formation, and symbolization. The secondary process operates in accordance with the reality principle; it is oriented toward the sought-for object in reality, using delays of impulse, detours, and experimental action in logical thought, until the suitable object and modes of action have been found. Obviously, the distinctions mentioned represent ideal polar positions; any specific behavior must be assessed in terms of its relative position with respect to a logical continuum or the particular admixture of primary and secondary process. There are all degrees of transition, reflected by shifting levels of psychic functioning.

---

[1] This paper is based upon the author's doctoral dissertation in partial fulfillment of the requirements for the doctorate degree, New York University. It was read in summarized form at the American Psychological Association, New York, September 1961.

[2] This term is explained and discussed below in Theoretical Considerations section.

Though primary and secondary process were first defined in connection with dream interpretations, they are also good working concepts for the study of schizophrenic ideation (cf. Rapaport, 1951). The formal, descriptive characteristics of the primary process bear a striking resemblance to the formal characteristics of schizophrenic thinking.

For purposes of the present study, primary process is inferred from (and defined by) communicated thought exhibiting any one of the following characteristics: (*a*) Inappropriate or unadaptive intrusion of libidinal or aggressive contents; (*b*) Condensation, symbolization, fragmentations, loose or fluid association, syncretic or autistic logic, logical contradictions, and impaired reality testing. The secondary process is similarly inferred from and defined by thinking that is faithful to logic and to external reality, that is purposive and organized for efficient goal-directed activity, and showing only appropriate drive-elements.

It is not rare in clinical testing to turn up evidences of primary process in the records of only moderately neurotic patients. What distinguishes these individuals from institutionalized psychotics is the greater success of their regulative strategies in handling the primary-process encroachment. One person, for example, might casually see a "threatening" figure on Card IV of the Rorschach, and point out with some amusement the incongruity of the large feet and the small arms. A person of this type might characteristically deal with primary process by permitting it into awareness in an ego-syntonic, modulated, and controlled fashion without experiencing emotional havoc. The less healthy person might see the same figure, find it "horrible" and quickly give up the card. For example, certain psychotic patients perceive their primary-process manifestations as ego-alien, unpleasant, and highly disruptive foreign intrusions upon the ego.

Thus, it is necessary not only to measure the relative predominance of primary- and secondary-process thinking, but also to determine the extent to which a person is master of, or is mastered by, primary-process elements in his thinking. The method of choice, then, must not only facilitate the emergence of identifiable primary-process material, but also provide for evaluating or scoring the responses. Recently, Holt and his co-workers (Holt, 1956; Holt & Havel, 1960) have systematized the application of the psychoanalytic theory of thinking to evaluation of Rorschach performance. Holt's (1959) manual provides a way of gauging both the amount of primary process that appears in Rorschach protocols and the *effectiveness* with which these manifestations are controlled. This procedure operationally defines the Rorschach variables used in the present study.

PROCEDURE

The subjects (*S*s) for this study consisted of 40 patients newly admitted to two Veterans Administration hospitals.[3] They were all males with a basic diagnosis of schizophrenic reaction. None had reached his 55th birthday by the date of entry into the study, there was no history of central nervous system disease, and none of them had had lobotomies. The patients were randomly assigned to one of two groups. The experimental group received chlorpromazine; the control group received a placebo, lactose. These drugs were packaged in individual bottles with only the patient's name printed on them as a means of identification. Instructions for administration were the same for each patient. Throughout the experiment, whether the patient received the drug or a placebo was not known to anyone connected with the experiment. The daily dose for each of the drugs was four 100-milligram pills. The two kinds of pills were identical in size, shape, color, and taste. Medication was provided for each patient entered into the study for a 3-month period.

In order to find out whether the medications used had any effect on the patients in this study, research data were obtained from two sources: (1) pre- and postdrug Rorschach protocols, and (2) pre- and postdrug evaluations on the Multidimensional Scale for Rating Psychiatric Patients (Lorr, 1953).

---

[3] The author was especially fortunate in gathering his data at two hospitals (Togus and Manhattan) where a maximum of harmony and consideration exist. His colleagues in nursing, social work, psychology, and psychiatry have in no small measure contributed much of whatever substance this paper may possess.

1. Rorschach data: The relevance and utility of the Rorschach test in assessing primary-process thought manifestations have been described by Holt and Havel (1960) in a progress report on the use of the method which Holt devised for scoring primary-process thinking as manifested in Rorschach performance.

All Rorschachs were administered by the author. The resulting 80 protocols were scored by doctoral students in clinical psychology who had no knowledge of any patient's drug or diagnostic status. Fifty-three records were scored independently by two persons who then met to adjust their differences and arrive at a consensus. Because of the time-consuming nature of the scoring procedure, 27 records were done by a single scorer. The following summary measures were derived:

(*a*) *Sum Level 1 of primary process:* The Rorschach manual distinguishes between material that is more direct, intense, raw, or blatant (scored "Level 1"), and relatively more civilized or socialized manifestations of the primary process (scored "Level 2"). Since the present study used a schizophrenic population, the Level 1 summary score helps to measure the extent to which the pressure of drive-ideation has overridden realistic and logical considerations. Sum Level 1 of primary process consists of the total number of responses having scorable Level 1 primary-process manifestations. Responses having one or more full primary-process instances were tallied as one, and responses having only weak scores were tallied as one-half. The uncorrected rater reliability for this score was .90.

(*b*) *Mean Defense Demand:* Holt (1959) defines defense demand as

> the degree to which the very nature of the underlying idea, or (in the case of the formal variables) the way it emerges, demands that some defensive and controlling measures be undertaken in order to make it a more socially acceptable communication [pp. 41–42].

Every response given a primary-process score was rated on a 6-point scale for its defense demand, that is, for the inherent or implicit "outrageousness" built into the content of the response or the way it emerged. This rating represents a finer differentiation of the Level 1—Level 2 dichotomy. Responses rated at the upper extreme, that is, 4, 5, or 6, were those with the greatest need for defense. The mean defense demand consists of the sum of the ratings for defense demand divided by the total number of responses that were given a primary-process score. The uncorrected rater reliability for this score was .74.

(*c*) *Mean Defense Effectiveness:* Every response given a primary-process score was rated on a 6-point scale for its degree of success in coping with the defense demand of the response, with 1 the most effective and 6 the most ineffective, control of the primary process. Mean defense effectiveness is the sum of the ratings for defense effectiveness divided by the total number of responses having a primary-process score. For the present study, because of the above concentration on the more extreme, Level 1, responses, mean defense effectiveness was computed only for responses containing Level 1 material. The uncorrected rater reliability for this score was .57.

TABLE 1

COMPARISONS ON THE LORR MORBIDITY SCORE

Variable	*M*	*SD*	*t*
Predrug versus postdrug	45.15 & 39.90	3.88*	3.93
Preplacebo versus postplacebo	49.95 & 49.70	3.28	.08
Predrug versus preplacebo	45.15 & 49.95	3.40	1.41
Postdrug versus postplacebo	29.90 & 49.70	3.98*	4.97

$^{*}p < .01$.

2. The Multidimensional Scale for rating psychiatric patients: the Lorr Scale consists of two sections.

(*a*) The *interview observations* were made by teams consisting of a clinical psychologist and a psychiatrist. The teams were responsible for all later interviewing and rating. The important consideration was that the same team should rate the same patients during the entire period of the study to insure the greatest reliability feasible.

(*b*) The *ward observation* was conducted by a ward nurse who cared for the patient. The ward nurse was provided with copies of the scale, and familiarized herself with its contents. Two days prior to a rating, the ward nurse was notified in order that she might have ample opportunity to observe the patients' behavior in the areas to be rated. Only one ward nurse was used for observations on any given ward.

RESULTS

The results of the study first of all provide additional weight for the published reports that chlorpromazine is helpful in altering the symptoms and behavior of schizophrenic patients (Table 1). The Lorr Morbidity scores leave little doubt that patients who took the drug showed a greater reduction in anxiety, agitation, and emotional lability than the amount occurring from placebo alone.

In Tables 2 and 3, these behavioral changes are confirmed by the Rorschach findings, particularly the appreciable improvement in the defense effectiveness. The experimental data suggest that while, for the most part, psychotic thinking as measured by the total amount of blatant, defense-demanding primary process was unaffected by chlorpromazine, there was a pronounced favorable shift in the patient's *attitude* toward the heretofore disturbing ideation and his manner of

TABLE 2

3 SUMMARY MEASURES OF PRIMARY PROCESS FROM THE RORSCHACH AND NUMBER OF RESPONSES: MEANS OF GROUPS BEFORE AND AFTER THORAZINE AND PLACEBO

Measure	Predrug ($N = 20$)	Pre-placebo ($N = 20$)	Postdrug ($N = 20$)	Post-placebo ($N = 20$)
Sum Level 1	4.4	3.95	3.77	3.42
Mean defense demand	2.68	2.40	2.30	2.38
Mean defense effectiveness of Level 1 primary process	3.31	2.97	2.32	2.80
No. responses (*R*)	17.15	16.25	16.95	17.9

handling it. In other words, test-retest comparison of the drug-versus-placebo Rorschach protocols shows that chlorpromazine had no reliable tendency to remove signs of psychosis in the formal or content aspects of primary-process thinking. Rather, the major effect of the drug seems to be a qualitative one. Intrinsic psychotic phenomena (e.g., autistic logic, arbitrary combinations, logical contradictions) were not abolished but there was a definite modification of their emotional consequences. When symptoms that make the patient feel the need to be defensive and afraid of his own impulses are controlled by a drug, there is a striking effect on communication and socialization. It may be that the patient slowly tests himself out in his environment, and when he is successful, feels himself to be a more adequate person.

Table 4 shows that within the treated group, amounts of independently rated clinical improvement were highly correlated with the degree of improvement in mean defense effectiveness, that is, the effectiveness with which the patient coped with the more extreme (Level 1) manifestations of primary process in his Rorschach responses. It is striking that the same level of relationship held true in the placebo group: patients "spontaneously improved" by clinical criteria did so by the test criterion of defense effectiveness also. It is interesting that the low positive correlation between Lorr scale improvement and changes in the other Rorschach measures attained some statistical significances when the two samples were combined. There was, therefore, a slight but reliable tendency for patients whose ward behavior improved, for whatever reason, to express less extreme primary-process material in their second Rorschachs. Finally, the fact that the correlations between behavioral change and Rorschach changes were quite similar for both the drug and placebo groups would seem in full accord with the view that chlorpromazine helps facilitate "spontaneous remission."

An analysis of the postdrug Rorschach records sheds some light on the precise nature of the changes that have taken place.

1. The first line of figures in Table 5 shows that there is a significant shift in the form accuracy and organization of the responses. The higher level of reality testing is reflected in the more logical, objective, and critical attitude that the patient maintains. Unconvincing blot interpretations tend to become articulated and elaborated upon until they are less conspicuously weak. Vague form responses diminish.

2. According to the second line in Table 5, chlorpromazine also seems to have a far-reaching effect on a person's total defensive operations. The remainder of the table breaks this finding down further:

(*a*) Whereas verbalization on the original Rorschach reflected an uncritical acceptance

TABLE 3

3 SUMMARY MEASURES OF PRIMARY PROCESS AND NO. OF RESPONSES: GROUP COMPARISONS BY *t*

Measure		Predrug versus postdrug	Preplacebo versus postplacebo	Predrug versus preplacebo	Postdrug versus postplacebo
Sum Level 1	*t*	.07	.55	.05	.18
Mean defense demand	*t*	1.9	.13	1.75	.49
Mean defense effectiveness of Level 1 primary process	*t*	2.77**	1.60	1.39	2.35*
No. responses (*R*)	*t*	.002	.016	.007	.019

* $p < .05$.
** $p < .01$.

TABLE 4

RANK-DIFFERENCE CORRELATIONS BETWEEN CHANGES AS MEASURED BY LORR SCORES AND BY 3 SUMMARY MEASURES (THORAZINE STUDY)

Group	Decrease in sum level 1	Decrease in defense demand	Increase in defense effectiveness
Drug group ($N = 20$) rho	.33	.21	.74**
Placebo group ($N = 20$) rho	.32	.32	.83**
All *S*s ($N = 40$) rho	.35*	.31*	.85**

* $p < .05$.
** $p < .01$.

of primary-process material, the second time around patients frequently relinquished or replaced their responses ("Two men having a bloody fight—no! It looks now like they are dancing!"), and on occasion, modified their percepts ("A bat with two heads or maybe this end is the head and the bottom is the tail"). These are examples of positive changes in *sequence* (see Table 5, line 3).

(*b*) There was an appreciable increase in the effectiveness with which drug patients used *context* in controlling primary-process material (see line 4). This fact bespeaks an improved ability to justify the deviant elements in a response so as to make it less estranged from group thinking. They made themes that would otherwise be quite shocking more appropriate by relating them to a custom, a cultural institution, or some other plausible context—aesthetic, intellectual, or humorous. The chlorpromazine patients in the current study seemed especially adept at selecting movies, television shows, and newspaper articles which safely accommodated their personal expressions of primary process ("This is a monster with big black hands and wings. It goes after people and sucks their blood. It was once a man, but now it came back to life as a monster. The reason I say that is because I saw a Bela Lugosi picture in the auditorium last week, where they had a monster that looked nearly the same way").

3. The part of the defense effectiveness variable that seemed to be most profoundly affected by chlorpromazine was the material elicited on the affect inquiry.[4] Unfortunately, the affects were not quantified or subjected to a statistical analysis. Consequently, the author's account should be taken as subjective and impressionistic. Nevertheless, after taking chlorpromazine, many patients definitely seemed less inclined to link concepts with personal experiences, and when they did, the experiences were more often benign. There were fewer expressions of inadequacy, dis-

---

[4] Because *S*'s affective response to what he saw in the blots turned out to be revealing of the effectiveness of his control and defense, Holt introduced an innovation in administration called "affect inquiry." This consists of asking *S*, after he has explained each response, how he felt about it, whether he found it pleasant, unpleasant, or indifferent to contemplate.

TABLE 5

DETAILED ANALYSIS OF CHANGES IN COMPONENTS OF DEFENSE EFFECTIVENESS BEFORE AND AFTER THORAZINE AND PLACEBO

Component of defense effectiveness	Predrug versus postdrug	Preplacebo versus postplacebo	Predrug versus preplacebo	Postdrug versus postplacebo
	(p values, sign tests)		(Mann-Whitney **U** tests)	
Positive changes in form level	.02	.05	**U** = 167	**U** = 120*
Overall positive changes in controls and defenses	.04	.19	**U** = 153	**U** = 112**
Positive changes in *sequence*	.02	.11	**U** = 133	**U** = 112**
Positive changes in *context*	.05	.16	**U** = 141	**U** = 120*
Positive changes in *remoteness*	.13	.15	**U** = 137	**U** = 143
Positive changes in *defenses*	.04	.08	**U** = 129	**U** = 117**

* $p < .025$.
** $p < .02$.

satisfaction, and perplexity. More percepts were called "pleasant." Emotional lability was at a minimum. The most frequent reaction to inquiry about feelings was a shrug of the shoulder, as if to say, "It's only a blot on the card. It's not real."

DISCUSSION

There seems to be considerable merit to the view that the effects of chlorpromazine upon openly hostile, anxious, and resistive patients initiate some degree of modification of the schizophrenic orientation. At the very least, the affective blunting that seems to take place makes most patients more amenable to institutional management. More important though, is the possibility that the diminution of anxiety permits a certain number of patients to integrate enough of the surrounding atmosphere to realize what is acceptable behavior and to selectively withhold their pathological symptoms. The temporary relief may serve as a beginning from which the patient may attain further improvement if given the opportunity to participate in various rehabilitative programs. It might be argued that any positive changes can be wholly accounted for by the lack of emotional involvement and that what might make for good hospital adjustment would not necessarily be conducive to adequate social functioning. While this may be a suitable explanation for the results obtained with some of the drug patients, it seems like a gross oversimplification for many others. Those individuals who improved upon their form accuracy, conceptual organization, and critical judgment were doing more than just being emotionally detached. Putting it in terms of primary and secondary process, the changes that were found when the various components of defense effectiveness were examined indicate an increase in the efficiency of secondary process working over of primary-process material. Of course, these results must be interpreted in terms of a decrease, but by no means a disappearance, of the predrug signs. Patients show varying degrees of change. Presumably, if a patient continued to maintain his gain and were retested some time later, we might sometimes find that there was an actual recession of the primary process. As yet, the literature contains no adequate follow-up studies regarding what takes place after drug treatment is discontinued.

Both the Lorr scale and the Rorschach manual proved their worth in the present study. The Lorr scale has enjoyed a wide reputation for being able to pick up behavioral improvement in patients. Its discriminatory power in illuminating the relative advantage of chlorpromazine versus a placebo reaffirms its usefulness. As for the primary-process manual, even granted its early stage of development, it is a tool with wide possibilities. If it does nothing more than increase the clinician's sensitivity to the subtle changes taking place under the effect of drugs, this would be sufficient reason to continue working with it. If by finding operational definitions for the psychoanalytic concept of primary and secondary processes, one can understand changes in patients within the context of a meaningful theoretical system, the rewards for continuing to invest in the manual will be even more fruitful.

REFERENCES

BELLAK, L. *Schizophrenia—a review of the syndrome.* New York: Logos Press, 1958.

BRENNER, C. *An elementary textbook of psychoanalysis.* New York: International Universities Press, 1955.

FREUD, S. The interpretation of dreams. *The basic writings of Sigmund Freud.* (Trans. A. A. Brill) New York: Modern Library, 1938.

HARTMANN, H. Technical implications of ego psychology. *Psychoanalytic Quarterly,* 1950, 20, 31–43.

HOLT, R. R. Gauging primary and secondary processes in Rorschach responses. *Journal of Projective Techniques,* 1956, 20, 14–25.

HOLT, R. R. Manual for the scoring of primary process manifestations in Rorschach responses. Draft 7, Research Center for Mental Health, New York University, April 1959.

HOLT, R. R., & HAVEL, J. A method for assessing primary and secondary process in the Rorschach. In M. A. Rickers-Ovsiankina (Ed.), *Rorschach psychology.* New York: Wiley, 1960. Pp. 263–315.

LORR, M. Multidimensional scale for rating psychiatric patients, hospital form. *Veterans Administration Technical Bulletin,* 10–507, 1953.

RAPAPORT, D. (Ed. & Trans.) *Organization and pathology of thought.* New York: Columbia University Press, 1951.

SCHAFER, R. Regression in the service of the ego: The relevance of a psychoanalytic concept for personality assessment. In G. Lindzey (Ed.), *Assessment of human motives.* New York: Rinehart, 1958. Pp. 179–191.

(Received March 22, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 253-255

FEAR AND AUTONOMIC AROUSAL[1]

JAMES H. GEER[2]

University of Pennsylvania

32 female undergraduates were classified as high fear (HF) or low fear (LF) of spiders. After reduction of orienting responses to pictures of neutral animals, Ss in the experimental condition were shown pictures of a spider, and Ss in the control condition were shown pictures of a snake. Using the response to the last neutral stimulus as a baseline, HF Ss in the experimental condition when first shown a spider yielded greater GSRs than other groups. This increase in GSR responding for HF experimental Ss was, for 2 more trials, maintained relative to Ss in the control condition. HF Ss in the experimental condition also yielded GSRs of greater duration than all other groups when first shown the spider stimulus.

An indication of both the intensity and object of fears was obtained from a large group of subjects. On the basis of those data, subjects were classified as possessing either high fear (HF) or low fear (LF) of a particular object. In the experimental condition, subjects from both HF and LF groups were exposed to pictures of the feared object. In the control condition, subjects from both fear levels were exposed to pictures of a commonly feared object but one that was not relevant to the subjects. It was predicted that HF subjects in the experimental condition would show increased autonomic arousal, as indexed by the GSR, when compared to LF subjects in the experimental condition and when compared with both HF and LF subjects in the control condition.

METHOD

Subjects

Undergraduates were administered the Fear Survey Schedule-II (Geer, 1965). On the basis of that test, the classification HF was given to subjects who rated their fear of spiders as either "Very Much" or "Terror" and their fear of snakes as "None." The classification of LF was given to those subjects who rated both their fear of spiders and of snakes as "None." Sixteen HF and 16 LF females were randomly selected as subjects.

Procedure

The subject was seated in an arm chair in a small, dimly lit, sound-deadened room with both the experimenter and recording equipment outside the room. GSR electrodes were attached to the first two fingers of subject's right hand. The subject was told to remain quiet while a series of animal pictures would be flashed on a screen placed directly in front of her.

A 15-minute rest period preceded the presentation of stimuli. A total of 10 black and white pictures of 2-seconds duration were shown to each subject. The first seven were "neutral" stimuli and consisted of pictures of a horse, a bluejay, a turtle, a rabbit, a squirrel, a raccoon, and a cow. These neutral stimuli were presented to permit habituation of "orienting responses" to the presentation of pictures. Interstimulus intervals varied from 45 seconds to 75 seconds with a mean of 60 seconds. The order of both neutral stimuli and intertrial intervals was varied randomly.

The last three stimuli were test stimuli. Half of the subjects, equally divided among HF and LF groups, saw a picture of a spider on Trials 8, 9, and 10: for the remaining subjects on Trials 8, 9, and 10, a picture of a snake was presented. A picture of a snake was used as it was felt that the control stimulus should be a negative object yet one that was not relevant to the subjects. That snakes are generally considered negative is suggested by Geer's (1965) data that college women rated fear of snakes considerably higher than fear of spiders.

GSRs were obtained from a Fels Dermohmmeter with the output written on an Esterling-Angus recorder. Beckman Biopotential Skin Electrodes were used with EKG-sol electrode paste. Stimuli were presented by a Kodak Carousel projector. The projector's light source was controlled by a timer so that stimulus duration could be accurately controlled. The projector's mechanism for changing slides was actuated approximately 20 seconds following stimu-

---

[1] This investigation was supported in part by a grant from the Erie County Mental Health Fund and in part by Public Health Service Research Grant MH-08905-01 from the National Institute of Mental Health.

[2] The author wishes to thank Sandra Jarka who collected the data of this study. Data for this study were collected while the author was at the Psychology Department of The State University of New York at Buffalo.

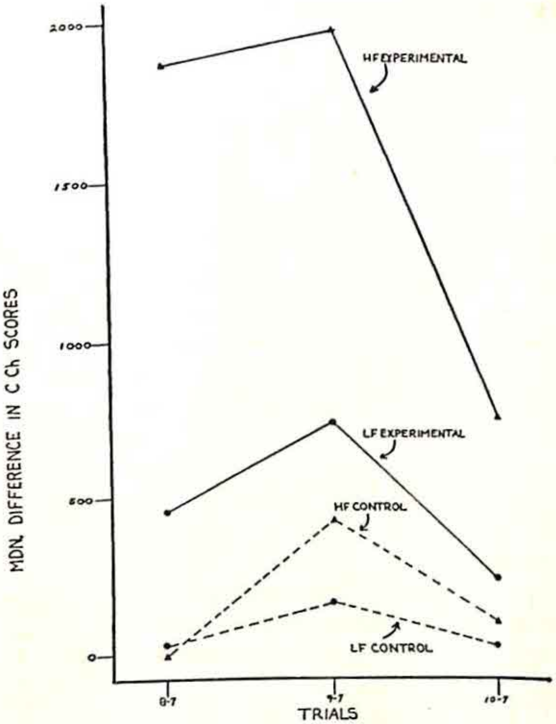


FIG. 1. Median difference in C Chs.

RESULTS

A GSR was defined as a decrease in skin resistance that began between 1 and 5 seconds following stimulus onset. GSR amplitude was measured by the conductance change score (C Ch) $(1/R_2 - 1/R_1) \times 10^9 + 1.0$ where R_1 was the resistance, in ohms, at the point of stimulus onset and R_2 was the resistance, in ohms, at the point of maximum GSR decrease during the 15 seconds following stimulus onset.

The dependent variable employed in assessing the amplitude of the response was the difference in C Chs for each subject between Trial 7, taken as a baseline, and subject's C Ch on the trial being considered. Figure 1 contains a plot of the Median values for differences in C Chs.

Since differences in C Chs were not normally distributed, Mann-Whitney U tests were used to assess group differences. On Trial 8, HF subjects in the experimental group yielded significantly greater ($p <.02$) C Ch differences than all other groups. On Trial 9, and Trial 10, HF experimental subjects yielded significantly greater ($p <.05$) C Ch differences than both control groups, and on Trial 10 the C Ch differences between HF and LF subjects in the experimental group just missed ($p <.08$) accepted levels of statistical significance. There were no significant differences among LF subjects in the experimental group and the two control groups. These results indicate that HF subjects, when first shown a stimulus relevant to their fear, yielded an increase in GSR responding. This increase, while tending to decrease over trials, continued larger for HF subjects in the experimental group than for subjects in the control groups.

The second GSR measure used was the time from the start of the GSR to the point of maximum change. HF subjects in the experimental group yielded significantly longer GSRs on Trial 8 than all other groups (Mann-Whitney U test, $p < .02$). On Trial 8, the duration measure did not differ significantly among the remaining groups. On subsequent trials the measure did not discriminate among groups.

DISCUSSION

The results of this study support the hypothesis that HF subjects will show increased autonomic arousal when shown stimuli relevant to their fear. These results were found with a picture of the fear object. This is of interest because it suggests that the results would have been even more pronounced had the actual feared stimuli been used. Fear, as expected, is selective. HF subjects did not yield increased GSR responding to a negative stimulus that was irrelevant to their fear.

The most plausible explanation of the increased GSR to pictures of feared stimuli is that the GSR is reflecting an increased emotional response in HF subjects. This is what would be expected since HF subjects were selected because they reported a negative emotional response to spiders. One alternative hypothesis should be noted. It may be that the increase in GSR responding is a reflection of differential orienting responses of the sub-

---

lus onset. The 20-second interval permitted full expression of any GSR to the experimental stimuli and also permitted the decay, prior to presentation of the next stimulus, of any GSR to the sound of the changing mechanism.

jects. It may be that HF subjects are responding with increased attention directed toward what for them is a significant stimulus. The present study cannot evaluate the relative merits of the alternative hypotheses.

REFERENCE

GEER, J. H. The development of a scale to measure fear. *Behaviour Research and Therapy*, 1965, 3, 45–53.

(Received April 23, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 256–259

PSYCHOPATHOLOGY AND SOCIOMETRIC CHOICE[1]

RUSSELL EISENMAN

Temple University

Previous research has suggested that except for paranoid schizophrenics, similarity in diagnostic category has little effect on sociometric choice among patients. A sociometric study was conducted employing MMPI scales as well as diagnostic categories on 84 hospitalized psychiatric patients. When diagnostic labels from the patients' charts were used, results showed no relationship between sociometric choice and diagnosis. However, patients classified by highest MMPI scale or by MMPI code type made positive sociometric choices among themselves.

Psychopathology has been related to difficulties in interpersonal relationships (Ansbacher & Ansbacher, 1956; Eisenman, 1965; Fromm-Reichmann, 1960; Garmezy, 1965; Sullivan, 1947). It is often hypothesized that the person labeled "mentally ill" withdraws from others because of feelings of inadequacy, anxiety, or fears of rejection. Such reasoning might lead to the formulation that the greater the mental illness, the greater the withdrawal. McMillan and Silverberg (1955) compared sociometric choices made on the following hospital wards: (*a*) neurological; (*b*) gastrointestinal medical; (*c*) open psychiatric, with neurotics and psychotics; (*d*) insulin therapy, with anxiety neurotics; and (*e*) closed psychiatric, with psychotics. There was limited support for the hypothesis of an increase in withdrawal as mental illness increased.

In an effort to extend the findings of the above study, Murray and Cohen (1959) compared sociometric choices on three wards: (*a*) a control medical ward, (*b*) an open psychiatric ward, and (*c*) a locked psychiatric ward. Sociometric choices of other patients were made on 12 items, with the findings that only paranoids tended to choose one another, and that social organization decreased as degree of mental illness increased.

The aspect of the Murray and Cohen (1959) study of concern here is their conclusion that sociometric choices were not made within diagnostic categories, except for paranoid schizophrenics. While this finding grew out of a study concerned primarily with degree of social organization on different wards, it is important because it suggests that the kind of patients with which the inmate associates have little, if anything, to do with his psychopathology, at least as far as diagnosis is concerned. Further, this same finding appears in a study by Shipman (1957) whose data show that, except for paranoids, sociometric choice is not related to diagnostic category.

Before concluding that people with various forms of psychopathology do not seek out others with the same form of psychopathology, more extensive research is needed. Diagnostic labels are notoriously unreliable, in part due to their reliance upon intuitive impressions of interview behavior or unvalidated tests. Consequently, the purpose of this study was to compare sociometric choices of patients at a mental hospital against (*a*) their diagnostic labels as revealed by their charts, and (*b*) their MMPI scores.

METHOD

A total of 84 patients in two wards at a state mental hospital were used. Half were on an intensive-treatment ward, while the other half were on a chronic ward. The sociometric questionnaire was the same as that employed by Murray and Cohen (1959) except for minor differences. The questionnaire was made up of the same kind of items that these investigators reported in their study, and an attempt was made to follow an identical procedure for selecting subjects (*Ss*). Specifically, the questionnaire was mimeographed in a booklet marked "confidential," with spaces for the patients to provide their name, age, and other identifying data. Patients were to be omitted if they could not provide this information. The *Ss* were not asked for their ward number as were the patients in the Murray and Cohen (1959) study, since few at the

---

[1] Numerous psychology technicians contributed to this research by administering and scoring tests. Their assistance is gratefully acknowledged.

hospital studied would know this. Patients were asked to make choices on the following 12 sociometric items. There were four positive items: like most, at ease with most, like to eat with most, like to work with most; four negative items: like least, ill-at-ease with, like to eat with least, like to work with least; and four neutral items: tallest patients, heaviest patients, most intelligent patients, and best-liked patients. As in the previous study, there were four spaces for names under each item, but patients were permitted to write in more names. The 12 sociometric items were in random order, with no effects found due to order of presentation of the item.

The official diagnosis was obtained from the patient's chart, and double-checked by investigating the diagnosis listed for each patient in the central record room. The MMPIs were usually already available, but where they were not they were administered.

The general nature of the questionnaire was explained to patients on both wards, and their confidence was obtained by telling them that the study was strictly confidential. Attempts were made to keep patients from talking to each other while filling out the sociometric questionnaire. Communication between patients was kept to a minimum. Those who could not or would not appear at the testing sessions were individually asked to take the test, and in all cases cooperation was obtained, without threats or inducement other than telling them it was for research purposes.

Social background information on each patient was obtained regarding the following: age, education, length of hospitalization, marital status, sex, and population area of origin. The patients were chosen in such a way that half were male and half were female. The overall breakdown was: chronic ward, 21 males, 21 females; intensive-treatment ward: 21 males, 21 females. In the hospital where this study was conducted, the male and female wards are separate.

RESULTS

Social background data are shown in Table 1, where it can be seen that there were only minimal differences between males and females, and between the chronic ward and the intensive-treatment ward. Analyses by chi-square tests indicated no statistically significant differences.

TABLE 1

SOCIAL BACKGROUND VARIABLES IN RELATION TO SEX AND CHRONIC OR INTENSIVE-TREATMENT WARD

Variable	Chronic ward		Intensive-treatment ward	
	Males ($N = 21$)	Females ($N = 21$)	Males ($N = 21$)	Females ($N = 21$)
Age (M)	35.5	37.6	34.1	33.4
Highest grade completed (M)	9.0	9.7	10.0	10.0
Length of hospitalization (M in years)	3.5	3.3	1.8	1.9
Married (N)	7	9	9	10
Population area of origin (N from towns under 10,000 population)	11	10	13	9

Table 2 reveals that there was no significant tendency for patients in the same diagnostic categories to choose each other on any of the three sociometric categories: positive, negative, or neutral items.

The MMPI results were analyzed in two ways: (*a*) by considering only the patient's highest scale score, and (*b*) by classifying the MMPIs by code type. Table 3 shows the number of patients with peaks on the various scales, and presence or absence of reciprocation of choices on the positive sociometric items. The four significant findings for patients with highest MMPI scale scores on *Hs*, *Pd*, *Pa*, and *Sc* indicate that these patients tended to choose each other on the positive sociometric items.

TABLE 2

NUMBER OF PATIENTS WITH ONE OR MORE RECIPROCAL CHOICES BY DIAGNOSTIC CATEGORY FOR POSITIVE, NEGATIVE, AND NEUTRAL SOCIOMETRIC ITEMS

Sociometric items	Diagnosis				
	Paranoid schizophrenic	Other schizophrenic	Character disorder	Neurotic	Other
Positive	11	6	7	5	1
Negative	6	6	5	7	1
Neutral	5	5	7	6	2

Note.—$\chi^2 = 3.75$, $df = 8$, $p > .80$.

TABLE 3

NUMBER OF PATIENTS WITH THEIR HIGHEST SCALE SCORE AND PRESENCE OR ABSENCE OF RECIPROCAL CHOICES ON POSITIVE SOCIOMETRIC ITEMS

Scale	Reciprocation	No reciprocation	x^2
L	0	0	—
F	2	1	0.33*
K	1	1	0.00*
Hs	10	2	5.32**
D	6	3	1.00*
Hy	1	0	1.00*
Pd	12	4	4.00**
Mf	2	1	0.33*
Pa	20	2	14.73***
Pt	0	1	1.00*
Sc	11	3	4.57**
Ma	0	1	1.00*

* $> .99\ p < .50$.
** $p < .05$.
*** $p < .01$.

MMPIs were coded using Hathaway and Meehl's (1951) method. The only significant findings which emerged were on the high-point codes (*t* scores above 54) in relation to the positive sociometric items. These results are summarized in Table 4.

None of the negative or neutral sociometric items showed a significant effect due to either diagnostic label or MMPI scale elevation or code type. However, number of choices and percentage of reciprocal choices made were significantly lower ($p < .05$) on the chronic ward suggesting that disturbed social relationships are related to more severe psychopathology (McMillan & Silverberg, 1955; Murray & Cohen, 1959). This inference rests on the assumption that patients on the chronic ward had more severe symptomatology.

TABLE 4

MMPI CODE TYPE AND PRESENCE OR ABSENCE OF RECIPROCAL CHOICES ON POSITIVE SOCIOMETRIC ITEMS

Code	Scales	Reciprocation	No reciprocation	x^2
489	*Pd, Sc, Ma*	7	1	4.50*
8′27	*Sc, D, Pt*	6	0	6.00*
4′9	*Pd, Ma*	9	1	6.40*

Note.—Prime (′) indicates *T* score above 70.
* $p < .05$.

DISCUSSION

The present investigation found reciprocation on positive sociometric items among patients with similar peaks and similar code types on the MMPI. When diagnostic labels were used for comparison with sociometric choices, no significant findings emerged. This suggests that the MMPI may be a better indicator of symptomatic similarity than the conventional diagnostic labels. Consequently, research which wishes to explore similarities or differences between subcategories of patients might do better to employ the MMPI or some related tool, instead of diagnostic labels. As the Zigler and Phillips (1961) study suggested, there is likely only a small relationship between diagnostic labels and specific symptoms.

Regarding the significant relationships between code types and positive sociometric choices, it may be that certain personalities tend to cluster together or seek each other out within the confines of the hospital. The 489 code type reflects what has been termed the "excitor scales," and has been linked to juvenile delinquency (Hathaway & Monachesi, 1953, 1957) and to high aggressiveness among eighth-grade boys (Butcher, 1965). These previous studies have not necessarily found their *S*s to have this same code type, but they have found their delinquent or aggressive *S*s to peak on these scales. The present study suggests that people who have the same code type may have similar personalities or similar behavior, and may tend to reciprocate positive sociometric choices.

REFERENCES

ANSBACHER, H. L., & ANSBACHER, R. R. (Eds.) *The individual psychology of Alfred Adler*. New York: Basic Books, 1956.

BUTCHER, J. N. Manifest aggression: MMPI correlates in normal boys. *Journal of Consulting Psychology*, 1965, **29**, 446–454.

EISENMAN, R. Usefulness of the concepts of inferiority feeling and life style with schizophrenics. *Journal of Individual Psychology*, 1965, **21**, 171–177.

FROMM-REICHMANN, F. *Principles of intensive psychotherapy*. Chicago: University of Chicago Press, 1960.

GARMEZY, N. Adaptive mechanisms in schizophrenia. *Bulletin of the Menninger Clinic*, 1965, **29**, 24–36.

HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: University of Minnesota Press, 1951.

HATHAWAY, S. R., & MONACHESI, E. D. *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: University of Minnesota Press, 1953.

HATHAWAY, S. R., & MONACHESI, E. D. The personality of pre-delinquent boys. *Journal of Criminal Law, Criminology, and Police Science*, 1957, **48**, 149–163.

MCMILLAN, J. J., & SILVERBERG, J. Sociometric choice patterns in hospital ward groups with varying degree of interpersonal disturbances. *Journal of Abnormal and Social Psychology*, 1955, **50**, 168–172.

MURRAY, E. J., & COHEN, M. Mental illness, milieu therapy, and social organization in ward groups. *Journal of Abnormal and Social Psychology*, 1959, **58**, 48–54.

SHIPMAN, W. G. Similarity of personality in the sociometric preferences of mental patients. *Journal of Clinical Psychology*, 1957, **13**, 292–294.

SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.

ZIGLER, E., & PHILLIPS, L. Psychiatric diagnosis and symptomatology. *Journal of Abnormal and Social Psychology*, 1961, **63**, 69–75.

(Received June 2, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 260–266

EYELID CONDITIONING AND AROUSAL IN SCHIZOPHRENIC AND NORMAL SUBJECTS[1]

BERNADETTE SPAIN[2]

Medical Research Council, Social Psychiatry Research Unit, Institute of Psychiatry, Maudsley Hospital, London

A group of schizophrenic patients were compared with normal Ss on an eyelid-conditioning task. All Ss received 50 trials to either a light or a tone CS and the UCS was an air puff to the eye. The principal findings were as follows: (a) the number of conditioned responses was shown to be related to level of skin potential in both normal and schizophrenic Ss, (b) the schizophrenic group gave more conditioned responses than the normal group and this was associated with higher skin potentials in the patient group, (c) the level of skin potential was found to be related to withdrawal and the more withdrawn Ss gave more conditioned responses, (d) a differential effect was found between normal and schizophrenic Ss and the modality of the CS.

Experiments on classical conditioning of schizophrenic patients do not lead to clear conclusions regarding these patients' ability to acquire conditioned responses. While most experimenters in the West can find no difference between patients and normal subjects (*S*s) (Franks, 1954; Howe, 1958; King & Landis, 1943; O'Connor & Rawnsley, 1959; Pfaffman & Schlosberg, 1939), Russian studies suggest that most schizophrenics tend to condition poorly (Lynn, 1963; Pavlov, 1941). In contrast, Spence and Taylor (1954) report acquisition levels higher than normal, and Mednick's (1958) theory of schizophrenia also predicts that at least acute patients should condition more rapidly than normal *S*s.

It is common to find that experiments on schizophrenic *S*s report conflicting results, but it is hardly surprising to find that groups of these patients do not behave uniformly. The diagnosis covers such a variety of symptoms that a limited number of patients used in any one study may differ in many essential characteristics from patients used in another study of the same phenomena. Since the conditioning experiments cited do not always relate the level of conditioning to symptomatology or other relevant characteristics of the patients, the discrepancy between results may well be due in part to these factors.

Work by Venables (1960, 1963a, 1963b) and Venables and Tizard (1958) suggests that schizophrenic reactions differ in many important ways according to the modality of the stimuli employed. It seems possible, therefore, that the modality of the conditioned stimulus (CS) might affect the acquisition of conditioned responses by schizophrenics and that this variable might also contribute to the disagreement among experiments on conditioning.

Many different responses have been used in studies of conditioning in schizophrenia. However, while there is some evidence that the same parameters may affect the conditioning of different responses (Razran, 1957; Spence, Haggard, & Ross, 1958), there is no experimental evidence for a factor of conditioning (Franks, 1958). Therefore, generalization from any one experiment on conditioning in schizophrenia must be limited, and it is quite possible that schizophrenics will condition poorly if one response is used, and rapidly if another response is chosen.

The following experiments attempt to overcome some of these difficulties. Certain characteristics of schizophrenic patients have appeared in previous work to be relevant to their performance in a variety of situations (Venables, 1960, 1963a, 1963b; Venables &

---

[1] This work was carried out in partial fulfillment of an MA degree, under the supervision of P. H. Venables, while the author was receiving a scholarship from the Medical Research Council. The experimental work was carried out at Springfield Hospital, with the kind permission of the Medical Superintendent, C. H. Beccle.

[2] The author is now working for the National Association for Mental Health.

O'Connor, 1959; Venables & Tizard, 1958). It seems possible that these variables might influence the acquisition of the conditioned response also. These characteristics are: (*a*) The degree of withdrawal, measured by a rating scale administered to nurses; (*b*) The presence of paranoid symptoms, also obtained from nurses' ratings; (*c*) The standing level of skin potential, which is thought to be a measure of arousal (Leiderman & Shapiro, 1964; Venables, 1963a, 1963b; Venables & Wing, 1962).

The eyelid response was selected for conditioning since eyelid conditioning may be regarded as a type of conditioned avoidance. It is possible that schizophrenic performance in this type of situation might relate to withdrawal behavior, particularly since Silverman (1963) concludes that schizophrenic performance improves in situations where "termination of the noxious stimulus is contingent upon the performance of the correct response." In order to examine the modality effect, both a visual and an auditory CS were employed.

METHOD

Apparatus

Eyeblinks were recorded by a pair of electrodes placed above and below the left eye, by a technique similar to electrooculography. When the lid moves over the eye, a potential change occurs which may be amplified and recorded on an Ediswan portable E.E.G. recorder. The CS was either a tone of 1000 cycles per second and 60 decibels in loudness, or an increase in brightness of a milk glass disc, from 7 to 700 millilamberts, subtending a visual angle of about 10 degrees. The CS was 1000 milliseconds in duration, and the onset of the unconditioned stimulus (UCS) occurred 500 milliseconds after the beginning of the CS. The UCS was a puff of air to the cornea lasting 160 milliseconds. The force of air measured at *S*'s eye was 1 gram. The air puff was delivered from a polythene tube, 1 millimeter in diameter, which rested in a block attached by a boom to a pair of earphones through which the auditory CS was presented. The block was adjustable in three dimensions so that the position of the tube in relation to the *S*'s eye could be individually adjusted. The puff was delivered by a Schrader valve with a very rapid action, and operated by a relay system. All stimuli were programed and the intertrial interval varied between 15 and 25 milliseconds with a mean of 20 milliseconds.

Skin potential was recorded from a pair of electrodes, one on the hypothenar eminence and one on the forearm, by the method described by Venables and Sayer (1963).

Procedure

Standing level of skin potential was recorded from each *S* after he had been resting in a dimly lighted room for about 4 minutes. Then the eye electrodes were attached and each *S* was given 50 trials with either the light or the tone CS. The *Ss* were instructed to press a key at the termination of the CS, since it had been found in a pilot experiment that this helped to reduce the incidence of voluntary responses and also ensured that *Ss* did attend to the CS. The *Ss* were told that before the termination of the CS they would receive a puff of air that would remind them that they must remember to press the key.

Ward ratings were obtained from the charge nurse of each male patient, giving details about the extent of the patient's withdrawal and evidence of delusional symptoms. These rating scales are described in Venables (1957) and Venables and O'Connor (1959).

Subjects

The *Ss* were 32 male schizophrenic patients between the ages of 28 and 53, with a mean of 40.3 years, all of whom were chronic, that is, had been in the hospital for at least 2 years, but who were not lobotomized or receiving ECT at the time of the experiment. Patients were conditioned to both CSs in a balanced design so that half were conditioned first to the tone and half to the light CS. The two testing sessions were separated by a period of 6–8 weeks. The most interesting result from the main experiment was replicated on 22 chronic female patients aged between 28 and 57, with a mean of 41 years. These patients were tested on one occasion only. Twenty-four normal *Ss*, 12 men and 12 women, were tested under the same conditions. These *Ss* were either nurses or hospital technicians or volunteer *Ss* from civil service departments. Their age range was similar to that of the patient *Ss*.

The records of an additional 10 male and 7 female patients and 6 normal *Ss* were rejected because *Ss* had a high blink rate or gave too many voluntary responses. In addition, 5 male and 3 female schizophrenics did not cooperate sufficiently for their records to be useful, for example, they did not press the key for the termination of the CS, or they shut their eyes for long periods during the testing session. The decision to eliminate an *S* from the study was made before analysis was attempted, with the aim of limiting statistical treatment to records which could be scored unambiguously.

Method of Scoring

All lid movements occurring between 200 and 500 milliseconds after the onset of the CS were considered to be anticipatory conditioned blinks, except for voluntary responses. Voluntary responses were identified on a criterion of trace shape, that is, those resembling instructed blink responses in a pilot study, the trace being characteristically smoother and rounder than for conditioned blinks, and of longer

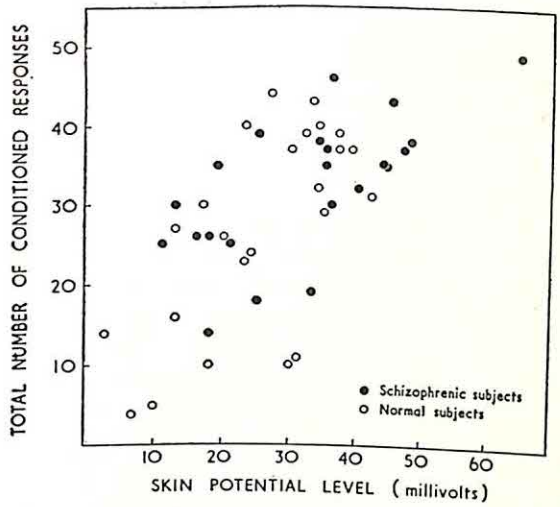


FIG. 1. The relationship between the level of skin potential and number of conditioned responses in female schizophrenic and all normal *S*s.

duration, lasting until after the onset of the UCS.[3] As a result of experience gained in the pilot study, it was decided to eliminate all records with a third or more anticipatory response of this type.

RESULTS

Conditioning Related to the Level of Skin Potential

A positive correlation was found between the number of conditioned responses and level of skin potential for all male *S*s from whom a negative skin potential could be obtained ($r = .44$, $N = 26$, $p < .05$). (For technical reasons skin potential could not be obtained from six *S*s.) Those *S*s with a high skin potential tended to give more conditioned responses, and this supports the hypothesis that level of conditioning is positively related to arousal. Correlations were obtained between age, chronicity, and level of skin potential, since, if it could be shown that level of skin potential is related to these variables, it would not be possible to infer that arousal influences conditioning. Both were nonsignificant ($r = -.29$ and $-.2$, respectively; $N\text{s} = 26$).

This finding was later confirmed in a repeat experiment using female patients. The correlation between skin potential and the number of conditioned responses was significant beyond the .001 level ($r = .66$, $N = 22$). With normal *S*s the correlation was very similar ($r = .63$, $N = 24$, $p < .001$; see Figure 1). It seems, therefore, that this relationship between level of skin potential and number of conditioned responses is not confined to schizophrenic patients, but is a more general characteristic.

Level of Conditioning

Figure 2 shows the increase in the number of conditioned responses made by the schizophrenic patients throughout the series of trials. Analysis of variance between the number of conditioned responses given in Trials 20–30 and Trials 40–50 confirms that conditioning did occur ($p < .01$). Furthermore, there was a significant increase in the number of responses between the first and second occasions of testing ($p < .01$; see Figure 2). Comparison between the number of conditioned responses given by normal and schizophrenic *S*s suggests that the latter tend to produce more conditioned responses than the normal *S*s ($t = 3.3$, $p < .01$; see Table 1). Further analysis of this result shows that this is associated with the fact that more patient *S*s show a high skin potential than do normal *S*s ($t = 2.6$, $p < .05$; see Table 2). When groups of normal and schizophrenic *S*s were

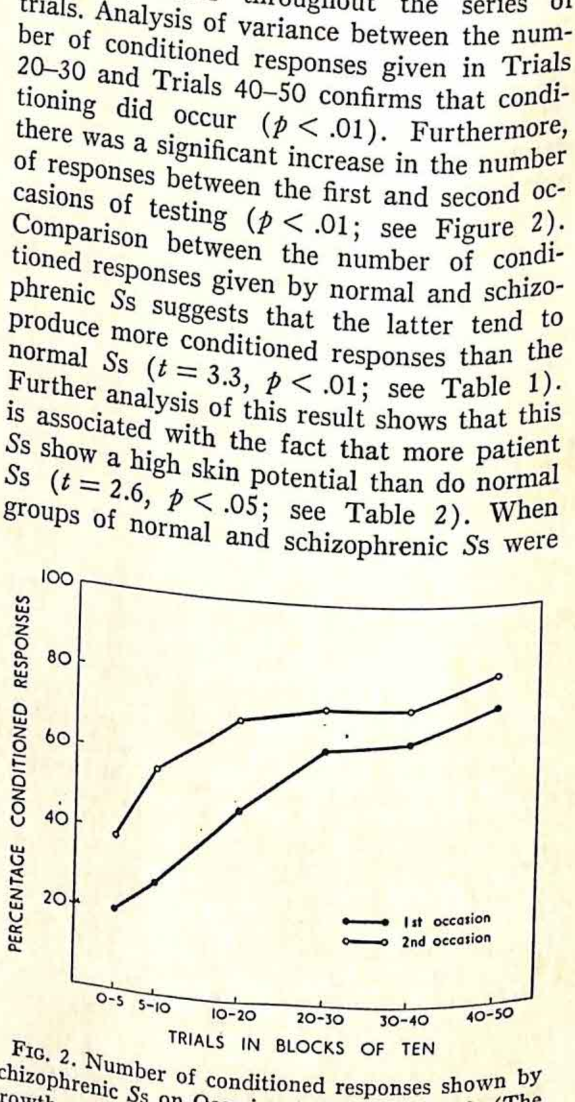


FIG. 2. Number of conditioned responses shown by schizophrenic *S*s on Occasion 1 and Occasion 2. (The growth curve of the responses from the normal *S*s was essentially similar.)

---

[3] Six records chosen because they presented difficulty in scoring were scanned, independently, by a colleague also engaged in eyelid-conditioning experiments. There was disagreement in only 10 out of 300 trials.

TABLE 1

NUMBER OF CONDITIONED RESPONSES FROM NORMAL AND PATIENT GROUPS

Group	*N*	*M*	*SD*	*t*
Normal	24	27.0	13.0	3.35*
Schizophrenic	48	36.6	8.3	

$* p < .01$.

TABLE 2

SKIN POTENTIAL LEVEL FROM NORMAL AND PATIENT GROUPS (IN MILLIVOLTS)

Group	*N*	*M*	*SD*	*t*
Normal	24	25.4	10.4	2.62*
Schizophrenic	48	33.0	13.5	

$* p < .05$.

matched on mean level of skin potential, there was no difference in the number of conditioned responses given by the two groups ($t = 1.6$, *ns;* see Table 3).

These results confirm the findings of Franks (1954) and O'Connor and Rawnsley (1959) that schizophrenics do condition, and does not support the Pavlovian hypothesis of pathological inertia in schizophrenic patients. The result partially supports Mednick's (1958) suggestion that schizophrenic patients condition better than normals because the former are functioning at a higher level of arousal.

Modality of CS and Number of Conditioned Responses

A comparison between the number of conditioned responses given by the patient groups to either a visual or an auditory CS, suggested that on each occasion of testing more responses appeared to the visual stimulus. A *t* test on a regression analysis for a group of *S*s matched on skin potential showed that significantly fewer responses were given to the auditory stimulus, when the effect of skin potential was held constant. The same trend does not appear in the data from normal *S*s, and a *t* test on a regression analysis of all the data from normal *S*s showed that these *S*s gave more responses to the auditory than to the visual stimulus; see Figure 3 and Tables 4 and 5. These two groups of normal *S*s did not differ in mean level of skin potential ($t = 1.6$, *ns*). The regression analysis was carried out on Trials 20–50, since, as can be seen in Figure 2, the asymptote of responding had approximately been reached by Trial 20.

TABLE 3

NUMBER OF CONDITIONED RESPONSES FROM NORMAL AND PATIENT GROUPS, MATCHED ON SKIN POTENTIAL LEVEL

Group	*N*	*M*	*SD*	*t*
Normal	24	32.3	12.6	1.60
Schizophrenic	24	27.0	10.2	

TABLE 4

PATIENT SUBJECTS: MEAN NUMBER OF CONDITIONED RESPONSES PER BLOCK OF 10 TRIALS (TRIALS 20–50 ONLY) BY MODALITY

Modality	*N*	*M*	*SD*	*t*
Light	80	7.6	2.39	3.16*
Sound	80	6.4	2.53	

Note.—Based on data from the regression analysis.
$* p < .01$.

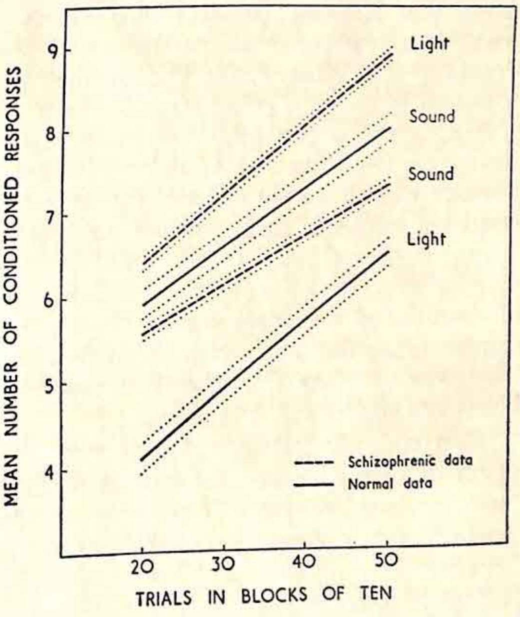


FIG. 3. Conditioned responses to auditory and visual stimuli by schizophrenic and normal *S*s, based on the results of the regression analysis. (The heavy lines represent mean scores and the dotted lines indicate the standard errors of the means.)

TABLE 5

NORMAL SUBJECTS: MEAN NUMBER OF CONDITIONED RESPONSES PER BLOCK OF 10 TRIALS (TRIALS 20–50 ONLY) BY MODALITY

Modality	*N*	*M*	*SD*	*t*
Light	44	5.3	3.09	2.66*
Sound	52	6.9	2.95	

Note.—Based on data from the regression analysis.
$* p < .01$.

Skin Potential, Symptoms, and Number of Conditioned Responses

Ratings of the degree of withdrawal were obtained on all male patients. Taking only those patients with no clear evidence of delusions, the correlation between skin potential and withdrawal was $-.47$ ($N = 19$, $p \leq .05$). This means that the more withdrawn patients gave the higher level of skin potential, which supports the hypothesis that the more withdrawn patients are more aroused than the active patients and confirms previous findings (Venables & Wing, 1962). Paranoid patients were excluded from this analysis since previous work by Venables and others has shown that paranoid patients differ in many ways from nonparanoid patients, and the Venables and Wing study showed that the relationship between skin potential and withdrawal is valid only for those patients who clearly are nondelusional. Withdrawal is not a function of either age or chronicity since correlations between these factors and withdrawal were not found to be significant. The correlation between withdrawal and number of conditioned responses was $-.32$, which is significant at the .05 level with an N of 32. The more withdrawn *S*s, therefore, tended to give more conditioned responses.

Thirteen male patients showed some evidence of delusions and the records of these patients were analyzed separately to see whether any differences could be detected between these patients and the nonparanoid patients, either in level of skin potential or number of conditioned responses. Only four of these patients showed a marked degree of paranoia. No significant differences appeared although there was some suggestion that the differential effect of modality of the CS was greater for the active paranoid patients.

All but 10 of the patients in this study were receiving some form of drug treatment and this factor was considered since it seemed possible that drug dosage might affect conditioning rates or arousal level. However, no systematic analysis could be attempted because of the great variety of drugs and drug combinations which these patients were receiving, and no significant relationship between drug status and any other variable could be determined, except that those patients not on drugs appeared to be more withdrawn than those on drugs.

DISCUSSION

It has been suggested that if *S*s' level of arousal increased throughout the conditioning session, this could produce an increase in their general responsiveness and consequently the increased number of blinks recorded in the experiment might represent a type of pseudoconditioning. The level of skin potential was measured only once, before conditioning commenced, so that it was not possible to determine in what way this variable was affected by the conditioning procedure. However, analysis of the number of blink responses occurring during two 5-minute intervals, one at the beginning and one at the end of the conditioning session, showed that there was no consistent rise in the periodic blink rate, that is, the general level of blinking, and that the increase in blinks recorded in the experiment was specific to the interval between the onset of the CS and the onset of the UCS. This would support the conclusion that true conditioning did occur.

These results, therefore, seem to indicate that schizophrenic *S*s are capable of acquiring a conditioned avoidance response at least as well as normal *S*s and that very withdrawn patients with a high skin potential condition better than normals. The present result is in keeping with previous findings on eyelid conditioning, although some investigators have reported no difference between normal and schizophrenic *S*s (Franks, 1954; O'Connor & Rawnsley, 1959). This is possibly because in these experiments the maximum possible variance was limited, since test trials were used, that is, conditioning was measured by counting responses only for the 18 trials on which the UCS was omitted.

It may be legitimate to regard eyelid conditioning as a type of conditioned avoidance, in that the conditioned blink can shield the eye, in part at least, from the effects of the air puff. In this case the relationship between withdrawal, skin potential, and level of conditioning would suggest the possibility that withdrawal itself represents a type of conditioned avoidance in patients who are overaroused and hypersensitive to noxious stimuli. This interpretation is implicit in the discussions of Bender and Schilder (1930), Garmezy (1952), Silverman (1963), and Chapman and McGhie (1963).

The relationship between conditioning and level of skin potential supports the findings of Runquist and Spence (1959) and Runquist and Ross (1959), who related the level of eyelid conditioning to measures of skin conductance, muscle-action potential, and heart rate. The result is congruent with the hypothesis that skin potential is a measure of arousal, as suggested by Venables (1963a, 1963b), particularly since the relationship was not confined to the patient population. It is not possible to predict from this experiment whether the same relationship would hold for other types of conditioned responses. An experiment by Fuster (1957) on the effect of direct stimulation of the reticular formation gives some basis to the prediction that arousal, at least if it is related to reticular arousal, might facilitate other types of conditioning.

The differential effect of the modality of the CS on conditioning in normal and schizophrenic *S*s is important in view of the finding on the differential effect of modality on reaction time (Venables & Tizard, 1958). While normal *S*s tend to give a shorter reaction time to auditory than to visual stimuli, schizophrenics give a shorter reaction time to visual stimuli. These facts, together with the present result, may indicate a difference between schizophrenics and normals in the arousing properties of visual as opposed to auditory stimuli.

REFERENCES

BENDER, L., & SCHILDER, P. Unconditioned and conditioned responses to pain in schizophrenia. *American Journal of Psychiatry*, 1930, **10**, 365–384.

CHAPMAN, J., & MCGHIE, A. A comparative study of disordered attention in schizophrenia. *Journal of Mental Science*, 1963, **108**, 487–499.

FRANKS, C. An experimental study of conditioning as related to mental subnormality. Unpublished doctoral dissertation, University of London, 1954.

FRANKS, C. Some fundamental problems in conditioning. *Acta Psychologica*, 1958, **14**, 223–246.

FUSTER, J. M. Tachistoscopic perception in monkeys. *Federation Proceedings*, 1957, **16**, 43.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *Journal of Personality*, 1952, **20**, 253–276.

HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic subjects. *Journal of Abnormal and Social Psychology*, 1958, **56**, 183–189.

KING, H., & LANDIS, C. Eyelid responses in normals and psychotic patients. *Journal of Experimental Psychology*, 1943, **33**, 210–215.

LEIDERMAN, H., & SHAPIRO, D. Galvanic skin potential level and some behavioural correlates. *Journal of Psychosomatic Research*, 1964, **7**, 277.

LYNN, R. Russian theory and research on schizophrenia. *Psychological Bulletin*, 1963, **60**, 486–498.

MEDNICK, S. A learning theory approach to research in schizophrenia. *Psychological Bulletin*, 1958, **55**, 316–327.

O'CONNOR, N., & RAWNSLEY, K. Two types of conditioning in psychotics and normals. *Journal of Abnormal and Social Psychology*, 1959, **58**, 157–161.

PAVLOV, I. *Lectures on conditioned reflexes.* (Trans. & ed. by W. H. Gantt) New York: International, 1941.

PFAFFMAN, C., & SCHLOSBERG, H. The conditioned knee jerk in psychotic and normal individuals. *Journal of Psychology*, 1936, **1**, 201–206.

RAZRAN, C. The dominance-contiguity theory of the acquisition of classical conditioning. *Psychological Bulletin*, 1957, **54**, 1–29.

RUNQUIST, W. N., & ROSS, L. The relation between physiological measures of emotionality and performance in eyelid conditioning. *Journal of Experimental Psychology*, 1959, **57**, 329–332.

RUNQUIST, W. N., & SPENCE, K. Performance in eyelid conditioning as a function of UCS duration. *Journal of Experimental Psychology*, 1959, **57**, 249–252.

SILVERMAN, J. Psychological deficit reduction in schizophrenia through response-contingent noxious reinforcement. *Psychological Reports*, 1963, **13**, 187–210 (*Monograph Supple.*, 2-V13).

SPENCE, K., HAGGARD, D., & ROSS, L. UCS intensity and the associative habit strength of the eyelid conditioned response. *Journal of Experimental Psychology*, 1958, **55**, 404–412.

SPENCE, K. W., & TAYLOR, J. A. Conditioning level in the behavior disorders. *Journal of Abnormal and Social Psychology,* 1954, **49**, 497–503.

VENABLES, P. H. A short scale for rating "activity-withdrawal" in schizophrenics. *Journal of Mental Science,* 1957, **103**, 197–199.

VENABLES, P. H. The effect of auditory and visual stimulation on the skin potential responses of schizophrenics. *Brain,* 1960, **85**, 77–92.

VENABLES, P. H. Selectivity of attention, withdrawal and cortical activation. *Archives of General Psychiatry,* 1963, **9**, 74–78. (a)

VENABLES, P. H. The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. *Journal of Psychiatric Research,* 1963, **1**, 279–287. (b)

VENABLES, P. H., & O'CONNOR, N. Reaction times to auditory and visual stimulation in schizophrenic and normal subjects. *Quarterly Journal of Experimental Psychology,* 1959, **11**, 175–179.

VENABLES, P., & SAYER, E. On the measurement of the level of skin potential. *British Journal of Psychology,* 1963, **54**, 251–260.

VENABLES, P. H., & TIZARD, J. Paradoxical effects of the reaction time of schizophrenics. *Journal of Abnormal and Social Psychology,* 1958, **53**, 220–224.

VENABLES, P. H., & WING, J. K. Level of arousal and the subclassification of schizophrenia. *Archives of General Psychiatry,* 1962, **7**, 114–119.

(Received June 2, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 267–274

EFFECTS OF GROUP PSYCHOTHERAPY WITH HIGH ACCURATE EMPATHY AND NONPOSSESSIVE WARMTH UPON FEMALE INSTITUTIONALIZED DELINQUENTS[1]

CHARLES B. TRUAX
Arkansas Rehabilitation Research and Training Center and University of Arkansas

DONALD G. WARGO
Johns Hopkins University Medical Center

AND LEON D. SILBER
University of Massachusetts

The study was undertaken to test the hypothesis that time-limited group psychotherapy, led by therapists who offer high accurate empathy and nonpossessive warmth, would produce significant change in delinquent behavior. The obtained results support this hypothesis. Significant superiority of the treated Ss, as compared with a control group, was demonstrated in percentage of time spent out of institution during a 1-year follow-up, as well as on the "C" scale of the Minnesota Counseling Inventory (MCI) which was specifically designed to differentiate delinquents from nondelinquents. A major mediating effect of therapy was to change self-concepts in a positive direction which increased the congruence between the self and the ideal self. Group psychotherapy also had significant positive effects upon attitudes toward parents and siblings.

The present study is aimed at determining the effects of group psychotherapy characterized by a relatively high level of therapist-offered accurate empathy and nonpossessive warmth. Constructive behavioral or personality change in institutionalized delinquent girls receiving such treatment is compared to changes in delinquent girls receiving no psychotherapy.

Accurate empathy is defined by the research scale as the therapist's communication of his moment-to-moment understanding of the patient's feelings, experiences, and perceptions from the patient's personal vantage point; at relatively high levels the therapist verbalizes feelings or experiences which are present but not in the patient's current field of awareness. Nonpossessive warmth is measured by the research scales as the therapist's communication to the patient of an unconditional warmth or acceptance characterized by a "caring for" or "prizing" of the patient and his feelings, experiences, and current verbalizations; at relatively high levels the only conditionality of the therapist's warmth is his urging that the patient talk of personally relevant material.

Research evidence, built upon the pioneering work of Whitehorn and Betz (1954), has suggested that psychotherapy as currently practiced may have both positive and negative effects, with an overall effect comparable to the changes that occur in no-therapy control groups. When therapists offer high levels of accurate empathic understanding (AE) and nonpossessive warmth (NPW), their patients tend to show significant positive personality and behavioral change beyond that shown in control groups; whereas, when AE and NPW are low, the patients show either no change or negative personality or behavioral change in comparison with control groups[2]

---

[1] This research was supported by Research Grant No. RD 906 from the Vocational Rehabilitation Administration, Department of Health, Education and Welfare and is part of a larger program headed by Charles B. Truax, Robert R. Carkhuff, and Donald G. Wargo. Appreciation is given to Edward P. Williams for assistance in the data analysis.

This research was made possible by the generous cooperation and help of the Kentucky Department of Child Welfare and Kentucky Village. In particular, the authors wish to thank Robert McClure, Superintendent of Kentucky Village and Richard J. Clendenen, Commissioner, Kentucky Department of Child Welfare.

[2] W. A. Dickenson and C. B. Truax. Group counseling with college underachievers: Comparison with a control group and relationship to empathy, warmth, and genuineness. Unpublished manuscript. University of Kentucky, 1965.

(Carkhuff & Truax, 1965; Rogers, 1962; Truax, 1961b, 1963; Truax & Carkhuff, 1963, 1964). These findings hold for counseling and hospitalized mental patient cases in both individual and group psychotherapy. The research findings of Butler (1960) and Cartwright (1961) are consistent with this two-factor effect of psychotherapy, as were the earlier findings by Barron and Leary (1955) and Cartwright and Vogel (1960), indicating significantly greater outcome variability for therapy than for control patients. Support for the positive relationship between the combination of accurate empathy and nonpossessive warmth on the one hand, and patient case outcome on the other, is indicated by the findings of a growing number of other research studies (Barrett-Lennard, 1962; Bergin & Solomon, 1963; Berzon, 1964; Cartwright & Lerner, 1963; Strupp, Wallach, Wogan, & Jenkins, 1963; Truax, 1961b; and Truax, Carkhuff, & Kodman, 1965).

This research, then, is an attempt to extend findings from prior research with hospitalized mental patients, neurotic outpatients, and college under-achievers to institutionalized juvenile delinquent girls. Although individual and group psychotherapy have seen increasing use with delinquent populations, the data from the Cambridge-Sommerville Youth Project (Teuber & Powers, 1953), which reported no effects of psychotherapy in comparing a large population of treated and control delinquents, raises a serious question about the efficacy of psychotherapy in producing ameliorative changes in delinquent populations. In the Teuber and Powers study, however, certain therapists or counselors tend to have a high frequency of poor relationships with delinquents in therapy, while others tend to have a low frequency of such poor relationships. It is suggested that the null effects in that study of psychotherapy were the result of lumping together therapists who offered high levels, moderate levels, and low levels of accurate empathy and nonpossessive warmth.

The present hypothesis is that the delinquent girls in group psychotherapy high in AE and NPW would show significant superiority over the girls serving as control subjects in terms of : (*a*) ability to get out and *stay out* of an institution during a 1-year follow-up, and (*b*) changes in delinquency-prone response as measured by the "C" scale of the Minnesota Counseling Inventory (MCI).

Additionally, it would be expected that any positive effect of group therapy upon these two measures of delinquency or delinquency-prone behavior would be mediated through changes in self-concepts, concepts of the ideal self, attitudes toward family and peers, and perhaps through positive changes in emotional stability or anxiety level. To throw light upon their potential mediating role, changes pre- to posttherapy in such factors were studied for delinquents in group therapy compared to delinquents serving as control subjects.

METHOD

Therapy and control conditions. The girls were randomly assigned to therapy ($N = 40$) or control ($N = 30$) conditions from a total population of 70 available female inmates at Kentucky Village. The Kentucky Village delinquent institution already had in progress two groups of girls who were in long-term "group work" group therapy, so that the available population did not include those girls which the institution judged as most likely to profit from a group-therapy experience. The remaining girls from the total population of 110 girls in the institution were unavailable because of conflicts in scheduling time.

The girls receiving group psychotherapy and the girls serving as controls received essentially identical treatment except that the girls receiving group psychotherapy were required to attend the group meetings twice weekly. The scheduling of the sessions in the afternoon overlapped with the timing of work assignments and "release" or free time in the cottages, so any advantage of getting off work early for the therapy girls was perhaps counterbalanced by their having to attend sessions during their free time. The therapists did not participate in discharge planning, which together with all administrative decisions about each girl both in therapy and control conditions was handled by one of the institutional social workers.

Four group-therapy groups of 10 girls each were formed and each group met for a total of 24 sessions with its therapist on a twice weekly schedule over a 3-month period. In view of the fact that the typical turnover in institutions precludes the use of long-term group therapy with the majority of the population, the therapy was time-limited so as to provide a more critical and practical test of the effectiveness of group psychotherapy. In addition, there is some evidence suggesting that time-limited psychotherapy is therapeutically as effective as long-term psychotherapy (Shlien, Mosak, & Dreikurs, 1960).

Subjects. The delinquent girls used in the present study ranged in age from 14 to 18 years, with a modal age of 16 years. Approximately one third of the sample had resided in urban areas (cities of more than 50,000 population), one third in depressed Appalachia rural areas, and one third in small towns. The population as a whole was characterized socioeconomically as lower-middle and upper-lower class, with only three girls coming from upper-middle or middle-middle socioeconomic families.

Therapists. The four groups were randomly assigned to the two therapists who participated in the research. Both therapists had participated previously as therapists in research on group psychotherapy with hospitalized mental patients (Truax & Wargo, 1965) so their levels of accurate empathy and nonpossessive warmth were available from the prior research. Of 16 therapists participating in that prior group-therapy research, the two therapists in the present study ranked highest on accurate empathy, and in the upper one third on nonpossessive warmth. It was therefore predicted that the girls receiving group psychotherapy in the present study would receive relatively high levels of AE and NPW.

Measures. The measure of institutionalization prior to therapy was the percentage of time that each girl spent out of the institution during the immediately preceding 1,061 days (which represented the greatest period of institutionalization among all the girls). The posttherapy measure of institutionalization used was the percentage of time each girl spent outside the institution during the 344 days subsequent to the initiation of group therapy.

A slightly modified version of the Butler-Haigh *Q* sort for Self and for Ideal (Butler & Haigh, 1954) was used as the basic measure of self-concepts, both pre- and posttherapy, employing scoring procedures identical to those used by Dymond (Rogers & Dymond, 1954). The specific measures were: (*a*) the self-adjustment score, which is obtained by imposing a true-false type of dichotomy on the sort; (*b*) the *z* transformation of correlations between the self sort and the "expert sort," which is a standard sort developed at the Chicago Counseling Center representing the expert's view of a most adequate adjusted self-concept; (*c*) the correlation *z* transformation of the self sort and the ideal sort, giving a score of self-ideal concept congruency; (*d*) the ideal-self adjustment score; and, (*e*) the *z* transformation of the correlation between the ideal-self sort and the expert sort.

The Minnesota Counseling Inventory provided measures of change in attitude or personality functioning. The subscales used were: (*a*) Family Relationships, (*b*) Social Relationships, and (*c*) Emotional Stability. Finally, the Anxiety Reaction Scale was used pre- and posttherapy to estimate the effects of group therapy on the anxiety levels of the delinquent girls.

The two most critical measures for a delinquent population would be the percentage of time out of the institution, and the Conformity scale of the MCI, which was developed specifically to differentiate between children who were well-adjusted and responsible and those who were delinquent, being in trouble both with the law and the school authorities.

RESULTS AND DISCUSSION

The present study of the effectiveness of group psychotherapy characterized by high levels of therapist accurate empathy and nonpossessive warmth was based upon the selection of two therapists who in the past had provided very high levels of AE and NPW to patients in group therapy. It was assumed that high levels of AE and NPW would be offered the delinquent girls in the present study. If, in fact, low levels of AE and NPW were provided in the present study, it should be recalled, then the prediction would be that psychotherapy would lead to poorer outcome than no psychotherapy. Prior evidence (Truax, 1961a) has suggested that the level of AE and NPW occurring in psychotherapy was determined to a large degree by the therapist so that the assumption was made that high levels would be offered in the present research groups. However, as a check, all sessions in the present study of group psychotherapy were recorded, and randomly selected samples from each interview were analyzed to ascertain the levels of conditions actually provided. The mean levels of accurate empathy, based on a 9-stage scale (Truax, 1961a) obtained by pooling the ratings of five judges each on two samples from each session were for the four groups: 4.8, 5.0, 5.1, and 5.1 (*M* of 5.0). With respect to nonpossessive warmth, the obtained values based on a 5-stage scale (Truax, 1963), pooling four judges each on two samples from each session of each group yielded mean values per group of 3.8, 3.7, 3.5, and 3.5, respectively (*M* of 3.6).

In terms of both absolute and relative values, these levels of AE and NPW are indeed high enough to justify the anticipation of positive effects. Data obtained from four therapists working in another juvenile delinquent institution yielded average values for group psychotherapy of 3.0 for accurate empathy and 2.6 for nonpossessive warmth.

TABLE 1

MEAN DATA FOR HOSPITALIZATION AND DELINQUENCY ("C") SCALE

	Percentage of time out of institution			Delinquency ("C") Scale	
	Therapy	Control		Therapy	Control
3-year pretherapy	82.0	78.0	Pretherapy	18.3	17.8
1-year follow-up	54.4	40.0	Posttherapy	17.0	18.3
			Change	−1.3	+0.5

EFFECTS OF HIGH AE AND NPW THERAPY UPON OUTCOME

Since the assignment of delinquent girls to time-limited group therapy and control conditions was randomly made, with no matching involved, the pretherapy mean values for therapy and control groups were not identical on each measure used. To control for any possible differences pretherapy and their effects upon *change* from pre- to posttherapy, analysis of covariance was used to test for differences in change from pre- to posttherapy between the therapy and control groups.

Comparisons btween therapy and control cases were made using analysis of covariance for the two most critical measures of outcome: (*a*) the percentage of time out of institution during the 1-year follow-up;[3] and (*b*) change from pre- to posttherapy on the "C" (delinquency) scale of the MCI.[4] The obtained *F* values for therapy versus control were 45.86 ($p < .001$) and 4.38 ($p < .05$), respectively. The findings then indicated a significant superiority in outcome for the delinquents receiving group therapy over those receiving only the usual rehabilitation without time-limited group therapy. The means pre- and posttherapy for time out of institution and for pre- and change (from pre- to post-) on the "C" scale of the MCI are shown in Table 1.

In terms of the absolute values on the conformity measure, the control girls showed an increase on the "C" scale suggesting greater irresponsibility, impulsiveness, rebelliousness, and self-centeredness, while the girls receiving group therapy showed greater understanding of the need for social organization, more responsible behavior, and less rebelliousness toward authority. However, according to the criterion groups of the MCI even those girls receiving group therapy would still be classified posttherapy as delinquency prone. Thus, while therapy had significant positive effects, the improvement was not so dramatic as to suggest that the girls would become responsible, nondelinquent students. It is possible, however, that continued group therapy for additional time periods would have additional positive effects.

It is sobering to note that the control group (which received the usual rehabilitation program provided by the institution) showed no tendency toward less delinquency-prone behavior. The evidence is that, over the 3-month period studied with the MCI, the effect of institutionalization was, if anything, to reinforce delinquency-prone responses rather than to reduce them.

Figure 1 presents the percentage of time out of institution from the initiation of the research for the therapy and control groups. As can be seen from the data, the superiority of the subjects receiving group psychotherapy begins shortly after the termination of 3 months of therapy and continues at about the same level throughout the 344 days of observation. The fact that the group-therapy population maintains about the same degree of superiority throughout the follow-up observation period would suggest that psychotherapy results in some degree of lasting behavior change.

---

[3] The percentage of scores were analyzed after an *arc sine* transformation.

[4] As in all analyses except percentage of time out of institution, the analysis of covariance between therapy and control groups was performed on the *change* scores, controlling for effects of initial status by use of the prescore. The mean posttherapy scores are presented for the interested reader.

Measures of change in self- and ideal concepts. In an attempt to throw further light upon the dynamics of the differential change in the girls receiving time-limited group therapy and those receiving control conditions, analysis of covariance[5] was performed on the data obtained from the self and ideal Q sorts, using the pretherapy scores as covariates in the analysis of *change* from pre- to posttherapy. These analyses resulted in F values for change scores of: (*a*) Q sort self- versus expert correlation, $F = 10.08$ ($p < .005$); (*b*) Q sort self-adjustment scores, $F = 17.77$ ($p < .005$); (*c*) Q sort self-ideal self-correlations, $F = 10.02$ ($p < .005$); (*d*) Q sort ideal-self adjustment scores, $F = 2.53$ ($p > .10$); and (*e*) Q sort ideal-self versus expert correlations, $F = 2.02$ ($p > .10$). The mean values are presented in Table 2.

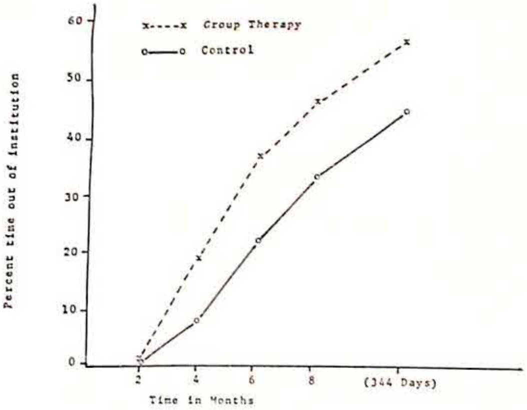


FIG. 1. Percentage of time out of institution for therapy and control female juvenile delinquents.

As can be seen, very little change in the self-concept measures occurred for the control girls. By contrast, there were large and statistically significant improvements seen in the girls receiving time-limited group therapy in terms of: (*a*) the adequacy of self-concept as measured by both the self-expert correlation and the self-adjustment score; and (*b*) the self-ideal concept congruency (as measured by the correlation between self and ideal). One major effect, then, of time-limited group therapy under high AE and NPW was improvement in the self-concept. Those girls receiving group therapy showed a significant shift in their self-conception toward a concept of the self more in line with "normal" girls and thus brought their self-concept into greater agreement with their ideal concept. Thus the increased congruence between self- and ideal concepts occurred primarily through shifts in self-concepts rather than in ideal concepts.

Changes in MCI personality measure. Three subscales from the MCI that would seem particularly appropriate for use with the delinquent adolescent girl are: family relationships, social relationships, and emotional stability. In most theoretical discussions and explanations of juvenile delinquency these three factors have been given prominent roles. The present data indicate that therapy had the effect of helping delinquent girls stay out of institutions during the 1-year follow-up, and of reducing delinquency-prone behavior as measured by the "C" scale: if these three factors indeed do play a significant mediating role in delinquent behavior, concomitant changes should occur in the measures of family relationships, social relationships, and emo-

TABLE 2

MEAN DATA FOR IDEAL CONCEPT AND SELF-CONCEPT Q SORTS

	Self-expert correlation		Self-adjustment		Self-ideal correlation		Ideal-adjustment		Ideal-expert correlation	
	Therapy	Control	Therapy	Control	Therapy	Control	Therapy	Control	Therapy	Control
Pretherapy	.38	.43	35.8	38.8	.44	.52	42.9	42.0	.57	.55
Posttherapy	.62	.39	44.3	36.6	.70	.44	46.1	41.9	.66	.56
Change	+.24	−.05	+8.5	−2.2	+.26	−.07	+3.2	−0.1	+.09	+.01

---

[5] The Q sort scores involving correlations per girl were transformed to z scores for analysis of covariance.

TABLE 3

MEAN DATA FOR MCI PERSONALITY MEASURES

	Family relationship scale		Social relationship scale		Emotional stability scale	
	Therapy	Control	Therapy	Control	Therapy	Control
Pretherapy	13.4	13.8	25.2	24.8	21.8	20.9
Posttherapy	10.6	14.3	19.1	19.0	16.6	18.1
Change	−2.8	+0.5	−6.1	−5.8	−5.2	−2.8

tional stability. As can be seen in Table 3, all differences, disregarding magnitudes, are in the predicted direction.

In comparison to standardization data from MCI both the therapy and control delinquents had moderately poor family relationships and felt their parents were unreasonably strict and demanding initially. After 3 months of group therapy the girls in the treatment groups changed to approximate very closely the normal population, while the controls showed no positive change ($F = 7.88$, $p < .01$). Those who had group therapy would be described posttherapy as having friendly and healthy relationships with parents and family; they usually regarded their parents as making reasonable demands on them and granting them a reasonable amount of independence and the girls looked forward to participating in activities with the family. That these improvements in family relationship scale scores for the therapy group were reflections of real attitudinal changes is perhaps best indicated by their better ability to go home and live successfully with their families during the 1-year follow-up.

In terms of the social relationship scale both groups showed marked and equivalent improvements (analysis of covariance F for therapy versus control of 0.96, $p > .40$). Initially, both groups of delinquent girls were within the range on the standardization data indicating poor social adjustment characteristic of students who seem to be unhappy and uncomfortable when with groups of students or adults. After 3 months, both the girls receiving group therapy and the control girls were well within the range of the normal population and could be characterized as relatively happy and comfortable in interacting with others; they appeared to enjoy talking with others, to have a genuine liking for others, adequate social skills, to converse easily, and conduct themselves appropriately in social situations. This finding suggests that the rehabilitation program of the institution and/or the enforced social proximity of the girls with each other, of itself, leads to better social relationships.

The emotional stability scale was designed to differentiate between stable and unstable students. Emotionally stable students are capable of making their own decisions, usually behave efficiently in emergencies, seldom worry, and tend to be calm and relaxed. Emotionally unstable students are frequently unhappy, over-react emotionally, lose their tempers easily, and in new situations are either fearful or overly aggressive. Initially, both the treatment and control delinquent girls were characterized by the normative data as showing quite poor emotional adjustment. Posttherapy, both groups showed improvement, particularly the therapy group, but there was no significant difference between therapy and control (F of 1.63, $p > .10$).

In summary, then, differences on all three scales of the MCI were in the predicted direction favoring greater improvement for the therapy than the control girls, but statistically significant differences occurred only on the family relationship scale.

Measures of change in anxiety levels. The role of anxiety in delinquent populations is little understood, but it provides converging information on emotionality. The anxiety reaction scale which has specific measures of: (*a*) general anxiety; (*b*) internalized anxiety (somatic expression of anxiety); and (*c*) social self-consciousness, was analyzed using

TABLE 4

MEAN DATA FOR ANXIETY REACTION SCALE

	General anxiety		Internalized anxiety		Social self-consciousness	
	Therapy	Control	Therapy	Control	Therapy	Control
Pretherapy	82.5	98.3	18.6	21.5	26.4	28.0
Posttherapy	63.7	83.7	17.3	20.8	20.4	24.5
Change	−18.8	−14.6	−1.3	−.7	−6.0	−3.5

analysis of covariance controlling for the initial levels and testing the significance of differences in change in anxiety levels between the control group and the treatment group. The means are presented in Table 4 where it can be seen that the differences between the treatment and control groups were in the predicted direction for the three anxiety measures but did not prove statistically significant. Both delinquent groups were initially lower in anxiety levels than typical outpatient groups, although higher than normal groups; posttherapy, both therapy and control girls, but particularly the therapy girls, approximated the normal population.

IMPLICATIONS

Present results indicate that group psychotherapy, characterized by relatively high levels of accurate empathy and nonpossessive warmth, produces positive changes in female juvenile delinquents beyond those observed in a control group. Greater positive change was observed in the therapy population than in the control population on each of the 13 measures used in the present study. That such greater changes for the therapy group are neither transient placebo effects (Frank, 1965) nor simply changes in test-taking behavior in an attempt to please the therapist is indicated by the *therapy group's greater ability to get out of the institution and stay out during the 1-year* follow-up.

The data suggest that effective psychotherapy with the delinquent girl is primarily mediated through: (*a*) changes in her self-concept toward greater self-valuing and a sense of adequacy and (*b*) changes in attitude about parents, seeing them as reasonable and sensible, which generalize to receptive rather than hostile attitudes toward authority figures.

Since "poor psychotherapy" as defined by low levels of AE and NPW was not studied, the findings are consistent with, but do not confirm the suggested explanation of the negative results of prior research on the effectiveness of psychotherapy with delinquents. That suggestion was that the erroneous assumption that psychotherapy was a unitary phenomenon led prior research to combine therapeutic, neutral, and destructive antecedents under the rubric of "psychotherapy." Confirmation of the suggestion that psychotherapy with delinquents may have positive, neutral, or negative effects depending upon whether the therapists offer high, medium, or low levels of AE and NPW will require further research. Such research should further clarify the present findings by comparing the effects of different levels of AE and NPW upon therapeutic outcome with institutionalized delinquents.

The present research does not contradict earlier research suggesting that "average psychotherapy" is ineffective with delinquents; it does indicate that "good psychotherapy," as defined by high levels of therapist-offered accurate empathic understanding and nonpossessive warmth, is both useful and effective in producing positive change in delinquents.

REFERENCES

BARRETT-LENNARD, G. T. Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs,* 1962, 76(43, Whole No. 562).

BARRON, F., & LEARY, T. Changes in psychoneurotic patients with and without psychotherapy. *Journal of Consulting Psychology,* 1955, 19, 239–245.

BERGIN, A. E., & SOLOMON, S. Personality and performance correlates of empathic understanding in psychotherapy. Paper read at American Psychological Association, Philadelphia, September 1963.

BERZON, B. The dimensions of interpersonal responsibility in therapeutic groups. *Western Behavioral Sciences Institute Report,* Special Series *I*, 1964.

BUTLER, J. M. Self-concept change in psychotherapy. *Counseling Center Discussion Papers,* 1960, **6**(13), 1–27. (University of Chicago)

BUTLER, J., & HAIGH, G. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. Rogers and R. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: University of Chicago Press, 1954.

CARKHUFF, R. R., & TRUAX, C. B. Training in counseling and psychotherapy: An evaluation of an integrated didactic and experiential approach. *Journal of Consulting Psychology,* 1965, **29**, 333–336.

CARTWRIGHT, R. D. The effects of psychotherapy: A replication and extension. *Journal of Consulting Psychology,* 1961, **25**, 376–382.

CARTWRIGHT, R. D., & LERNER, B. Empathy, need to change, and improvement with psychotherapy. *Journal of Consulting Psychology,* 1963, **27**, 138–144.

CARTWRIGHT, R. D., & VOGEL, J. L. A comparison of changes in psychoneurotic patients during matched periods of therapy and no therapy. *Journal of Consulting Psychology,* 1960, **24**, 121–127.

FRANK, J. D. The role of hope in psychotherapy. Paper presented at Kentucky Centennial Psychotherapy Symposium and Workshop, University of Kentucky, April 1965.

ROGERS, C. R. The interpersonal relationship: The core of guidance. *Harvard Educational Review,* 1962, **32**, 416–429.

ROGERS, C. R., & DYMOND, R. D. *Psychotherapy and personality change: Coordinated research studies in the client-centered approach.* Chicago: University of Chicago Press, 1954.

SHLIEN, J. M., MOSAK, H. H., & DREIKURS, R. Effect of time limits: A comparison of client-centered and Adlerian psychotherapy. *American Psychologist,* 1960, **15**, 415. (Abstract)

STRUPP, H. H., WALLACH, M. S., WOGAN, M., & JENKINS, J. W. Psychotherapists assessments of former patients. *Journal of Nervous and Mental Disorders,* 1963, **137**, 222–230.

TEUBER, H. L., & POWERS, E. Evaluating therapy in a delinquency prevention program. *Proceedings: Association for Research in Nervous and Mental Disorders,* 1953, **31**, 138–147.

TRUAX, C. B. The process of group psychotherapy. *Psychological Monographs,* 1961, **75**(14, Whole No. 511). (a)

TRUAX, C. B. Therapeutic conditions. *Discussion papers, Wisconsin Psychiatric Institute,* 1961, No. 13. (University of Wisconsin) (b)

TRUAX, C. B. Effective ingredients in psychotherapy: An approach to unraveling the patient-therapist interaction. *Journal of Counseling Psychology,* 1963, **10**, 256–263.

TRUAX, C. B., & CARKHUFF, R. R. For better or for worse: The process of psychotherapeutic personality change. Chapter in: *Recent Advances in the Study of Behavioral Change.* Montreal: McGill University Press, 1963.

TRUAX, C. B., & CARKHUFF, R. R. Significant developments in psychotherapy research. In Abt and Riess (Eds.), *Progress in clinical psychology.* New York: Grune & Stratton, 1964. Pp. 124–155.

TRUAX, C. B., CARKHUFF, R. R., & KODMAN, F., JR. Relationships between therapist-offered conditions and patient change in group psychotherapy. *Journal of Clinical Psychology,* 1965, **21**, 327–329.

TRUAX, C. B., & WARGO, D. G. Therapeutic conditions and patient personality change during group psychotherapy: Hospitalized mental patients. Unpublished manuscript, Kentucky Mental Health Institute, 1965.

WHITEHORN, J. C., & BETZ, B. J. A study of psychotherapeutic relationships between physicians and schizophrenic patients. *American Journal of Psychiatry,* 1954, **3**, 321–331.

(Received June 23, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 275–279

STABILITY OF SELF-CONCEPT IN SCHIZOPHRENIA

IRVING B. GULLER[1]

Franklin and Marshall College

This study was to investigate the stability of self-concept in nonchronic, nonparanoid schizophrenics. Self-Concept, Health-Concept, and Food-Preference questionnaires were administered to 84 Ss. Schizophrenics and controls were divided into failure and nonfailure groups ($N = 21$ per group). Test-retest method with failure or neutral condition immediately preceding the 2nd testing showed schizophrenics ($N = 42$) more variable for self-concept but not for food or health responses. Failure significantly increased self-concept variability for schizophrenics but not for controls. Schizophrenia appears to have a self-concept disorder component which manifests itself through inconsistent self-descriptions. It is suggested that some schizophrenic symptomatology may reflect efforts to establish a stable, if unrealistic, concept of self.

Sullivan (1953), Arieti (1955), and others have observed an association between schizophrenia and self-concept disorder. The most consistent finding in studies of self-concept in schizophrenia is a wide range of intragroup variability. Self-esteem or ideal-self congruence measures do not distinguish schizophrenics from normals. Self-concept stability in schizophrenia or the conditions affecting it have attracted little research attention. Rogers (1951) and Combs and Snygg (1959) suggested that self-concept disorder characterizes maladjusted individuals. Support for this view was offered by Brownfain (1952) and Engle (1959). They specifically found self-concept variability to be associated with poor social and emotional adjustment.

Failure is regarded as having an impact upon the self-concept requiring some reassessment of it (Diller, 1954; Katz, 1950). The less stable the self-concept organization, the more widespread the effects of failure should be. If schizophrenics have less integrated self-concepts than normals, failure should bring this deficiency into sharper focus.

Variability of self-concept responses might only reflect the generalized inconstancy of behavior often attributed to schizophrenics. But the evaluation of his interpersonal self might represent a singular problem for the nonchronic schizophrenic. Other material such as ideas about one's general physical condition and certain preference hierarchies bear less directly upon one's interactions with others and are less dependent upon the latter for their evolution. They should be more reliable and less vulnerable to failure than the self-concept.

For the purpose of this study, self-concept was defined as an organized system of relatively stable self-perceptions largely derived from and pertinent to interpersonal relationships. It was operationally defined as subject's responses to a questionnaire composed of self-referent statements. Stability was defined as consistency in response to the same items over time.

The hypotheses of this study were as follows:

1. Nonparanoid schizophrenics without evidence of deterioration or crystallized delusional thinking will show less stable self-concepts than normals matched for age, sex, intelligence, conceptual ability, and tendency to report honestly about themselves.

2. Schizophrenics will show more instability of self-concept than normals following induced failure, while other concepts similarly measured will be less affected.

3. Self-concept instability in schizophrenics is not simply an aspect of generalized variability in such patients. Less interpersonal

---

[1] This article is based upon an unpublished doctoral dissertation completed at New York University in 1962. The author wishes to thank Margaret Tresselt, research advisor, Robert Morrow, Chief Psychologist of the Bronx VA Hospital and Howard White, Chief Clinical Psychologist of the VA Center, Togus, Maine for their cooperation and recommendations.

TABLE 1

SAMPLE ITEMS FROM THE SELF-CONCEPT AND HEALTH-CONCEPT INSTRUMENTS

Item No.	Self-concept	Health-concept
4	I am strong willed	I have strong arms
6	I make enemies easily	I catch cold easily
23	I enjoy an intellectual challenge	I like to do physical exercise
44	My feelings are easily hurt	I am very sensitive to pain

material such as food preferences and concepts about one's typical physical condition will not show similar inconstancy when measured in the same manner.

METHOD

Subjects. 84 male subjects between the ages of 21 and 50 were divided into four groups of 21 each: failure schizophrenics (SF), nonfailure schizophrenics (SNF), failure controls (CF), and nonfailure controls (CNF). The subjects had completed between 8 and 14 years of education in the United States, were of at least low average intelligence, and had no suspected neurological disorders.

SF and SNF subjects were hospitalized schizophrenic veterans, in hospital less than 4 months, and without evidence of crystallized delusional thinking or deterioration. They had not been under psychiatric treatment for at least 6 months prior to the current admission. None were receiving convulsive therapy or high levels of medication (as judged by medical staff) at the time of, or just prior to, the study.

CF and CNF groups included staff and some general hospital patients about to be discharged from the same veterans' facilities.[2] Control subjects were without evidence of emotional disorder, psychiatric histories, serious illness, or records of frequent hospitalization. The general hospital patients employed were essentially recovered from minor illnesses. Potential subjects with recent serious illness or chronic conditions were not employed. This was to avoid the problems which could have arisen if sick or very somatically concerned people were dealing with the "Physical Health Concept" questionnaire (see instruments, below). Twelve were hospital aides, and several others were clerical staff. Groups were equated for age, education, intelligence, conceptual ability, and tendency to report honestly about themselves. The subjects were randomly assigned to failure and nonfailure groups.

Instruments. The WAIS Similarities subtest, Shipley Institute of Living Scale (Vocabulary), and a questionnaire composed mainly of "Lie" items from the MMPI, were used to equate groups for conceptual ability, intelligence, and honesty, respectively. Cutoff points were used to eliminate unsuitable subjects.

Three special questionnaires of 46 items each were prepared for this experiment. For the Health- and Self-Concept scales, 120 items of each content type were first created. The experimenter and two clinical psychologist judges paired items from the two lists on the criteria of length, clarity, complexity, and relevance to enduring types of self-assessments rather than those which might undergo changes from day to day or week to week with moderate changes in the environment or physical condition of the subject. The 46 pairs used to construct the final instruments were those all judges felt represented a close approximation of equivalence considering the disparate contents. The food scale involved preference ratings for 46 common foods. Items of the three scales each had eight scoring categories ranging from "Completely True" (designated 1) to "Completely False" (8), with intermediate ratings numbered accordingly. Instructions, administration, and scoring were consistent for the three tests. Instructions emphasized that the subject should describe himself the way he felt he usually was. Table 1 gives sample items from the Self-Concept and Health-Concept scales.

Failure task. Failure was induced using a modified Stroop (1935) color word interference task of 100 items. The subject was required to rapidly name the *color of ink* used to print four different color nouns. The latter were randomly arranged and printed in the four colors of the nouns. RED might be printed in green and BLACK in blue. The task was presented as "This will show something important about you" with no further explanation. All subjects hesitated while performing. They were stopped at a predetermined point (Item 47) and then shown their "score" within a failure category of bogus norms. The neutral task was an innocuous 3-minute chat with the experimenter.

Procedure. The subjects were administered the screening tests and the three special questionnaires during the first session. Testing was conducted in

TABLE 2

MEAN VALUES OF VARIABLES USED TO EQUATE GROUPS

Group	Age	Years education	Raw Shipley	Raw Lie	Raw Similarities
SF					
M	34.6	10.7	28.0	3.0	14.7
SD	6.9	1.3	3.6	1.7	3.9
SNF					
M	35.3	11.0	29.7	2.4	15.3
SD	7.5	1.8	5.0	1.7	2.7
CF					
M	34.7	11.1	27.6	3.0	14.3
SD	9.1	.6	4.5	1.6	2.8
CNF					
M	34.1	10.6	29.1	2.1	14.8
SD	8.9	2.0	4.4	1.6	3.4

---

[2] Veterans Administration Center, Togus, Maine and Veterans Administration Hospital, Bronx, New York.

groups of from two to five except with the WAIS Similarities which was individually given. The subjects were recalled 48 hours later and depending upon condition group, were given the failure or interview, immediately followed by the three questionnaires in the original order. They were then given a brief review of the experiment but without reference to the bogus nature of the failure. The subjects were also asked not to discuss the experiment with others and to the knowledge of the experimenter this request was honored. Finally, failure subjects were given a brief standardized interview to determine subjective reactions to the failure. Responses were tabulated for later analysis.

RESULTS

Groups were well equated for age, education, verbal intelligence, and conceptual ability. Table 2 shows the mean values of the data used to equate groups.

Lie score for SF and CF groups combined were higher than for the nonfailure subjects. The mean difference between means of condition groups was .75 ($F = 4.45$, $p \leq .05$). Though statistically significant, this difference is regarded as unimportant except that it may suggest some minor attenuation of variability (especially in self-concept) for the failure groups due to greater orientation toward socially approved responses.

Groups did not differ significantly with respect to tendency toward end responses on the experimental tests and, therefore, absolute per item difference scores were used to evaluate the hypotheses. These scores were determined by comparing a subject's first and second responses to each item. Thus, if an item was first answered "Mostly False" (designated 6) and on second testing "Completely False" (8), the net change (without regard to direction) would be 2. Such shifts were then totaled to arrive at subject's total difference score (TDS) for a particular instrument. Table 3 gives the TDS totals, means, and standard deviations for groups by instruments.

TABLE 3

TOTAL DIFFERENCE SCORES BY GROUPS AND INSTRUMENTS

Group	Self-concept	Health-concept	Food preference
SF			
Total	1,214	876	461
M	57.8	41.7	22.0
SD	17.9	10.1	11.3
SNF			
Total	916	839	538
M	43.6	39.9	25.6
SD	11.3	15.4	12.2
CF			
Total	897	792	501
M	42.7	37.7	23.9
SD	12.9	13.9	11.9
CNF			
Total	814	850	473
M	38.8	40.5	22.5
SD	12.9	15.5	10.3

TABLE 4

OVERALL ANALYSIS OF VARIANCE FOR ALL TDS SCORES

Source of variation	*SS*			*df*	*MS*	*F*
Between groups	1,611.25		3		537.08	1.83
Between diagnosis		1,060.68		1	1,060.68	3.61
Between failure		383.81		1	383.81	1.31
Diagnosis × Failure		166.76		1	167.76	----
Subtotals		1,611.25		3		
Subjects variation within groups (Error 1)	23,480.19		80		293.50	
Total between subjects	25,091.44		83			
Between contents	22,377.52		2		11,188.76	88.82*
Contents × Groups	3,100.56		6		516.76	4.10*
Contents × Diagnosis		1,100.23		2	550.11	4.37*
Contents × Failure		1,378.13		2	689.06	5.47*
Contents × Diagnosis × Failure		622.20		2	311.10	2.47*
Subtotals		3,100.56		6		
Content × Subject within group variation (Error 2)	20,154.59		160		125.97	
Total within subjects	45,632.67		168			
Totals	70,724.11		251			

* $p = <.05$ (for two-tailed test).

TABLE 5

2 × 2 ANALYSIS OF VARIANCE REPLICATED OVER 3 TESTS

Summary of F results	Food	Health	Self
Source			
Between diagnosis	—	—	10.16*
Between failure	—	—	8.40*
Diagnosis × Failure	—	—	2.67
Between failure groups (SF-CF)	—	1.12	9.34*
Between nonfailure groups (SNF-CNF)	—	—	1.59
Between schizophrenic groups (SF-SNF)	—	—	8.99*
Between nonschizophrenic groups (CF-CNF)	—	—	—

* $p < .05$ (two-tailed test).

TDS data were analyzed by two basic methods. The first was a three-way analysis of variance for repeated measures of the same population (Edwards, 1950). Results of this analysis are given in Table 4. The second method was a 2 × 2 factorial analysis of variance replicated over three tests. Results are summarized in Table 5. A significance criterion of .05 for two-tailed tests was applied throughout.

Schizophrenics ($N = 42$) were significantly more variable than controls in response to self-concept items (Table 4, Content × Diagnosis). This is consistent with the first experimental hypothesis. Group SF was significantly more variable for self-concept than were CF or SNF groups. Normal groups, however, did not significantly differ in this regard. Accordingly, the second hypothesis is regarded as confirmed.

Under neither treatment condition did variability of Health-Concept or Food-Preference responses differentiate diagnostic groups. Nor did failure significantly influence response variability of schizophrenics for these contents. These findings confirmed the third hypothesis of this study.

The SNF group was not significantly more variable for Self than for Health items though the trend was as predicted (Content × Diagnosis × Failure interaction, see Table 4 and between NF groups, see Table 5).

In the posttest interview, both diagnostic groups tended to minimize the impact which failure had upon them. All subjects appeared to believe that the failure was genuine and groups did not differ in reported motivation to do well.

DISCUSSION

The first hypothesis of this study was supported. Schizophrenics appear to have less reliable self-concepts than normals. A stress situation brought this fact into clearer focus. The size of NF groups and the fact that two-tailed tests were used, required greater absolute differences to reach significance. Self-concept stability of schizophrenics was significantly more disrupted by failure than was that of normals, but Food-Preference and Health-Concept responses were not similarly affected. If the schizophrenic's self-concept variability were simply a function of generalized response variability, the latter should have manifested itself equally on the Health-Concept and Food-Preference instruments. This did not occur. Self-concept seems to represent a special area of confusion for the schizophrenic.

It is possible that failure had more impact upon schizophrenics than controls because SF subjects were hospitalized for "mental trouble." They might have interpreted failure as a black mark against their desire for discharge or as confirmation that they were sick. Posttest interview data do not, however, support this notion. Brownfain noted that the less stable the self-concept, the greater tends to be one's dependence upon how the environment evaluates him. Higher self-concept variability in the SF group is regarded as reflecting the poorly integrated self-concepts which had to accommodate the failure.

The self-concept includes ideas of one's social acceptability and capacities. It may come under scrutiny by others (i.e., is partially open to consensual validation) and can be challenged by the experiences of life. A stable and reality oriented self-concept would appear important for effective social functioning.

Sullivan has observed the frequent progression in schizophrenia from a nonchronic state, typical of the SF and SNF subjects of this study, to a paranoid resolution involving fixed false beliefs. Some of the schizophrenics in this study were beginning to show early signs of paranoid development with suspiciousness and vague ideas of reference. Having established a factor of self-concept instability in the schizophrenic groups of this study, it is useful to question its potential significance. It seems possible that the progression toward paranoid states, as described by Sullivan and observable in the present schizophrenic group, can be interpreted from the standpoint of self-concept theory. Without a stable self-concept, one is apt to be constantly anxious, unable to plan ahead and have difficulty in communicating effectively or anticipating the reactions of others (Arieti, 1955; Lecky, 1945). By seizing upon some beliefs, however false, and holding to them tenaciously, the person becomes temporarily more secure. He may be unable to function outside of the protective environment of a hospital, but this is a small price to pay for relief from anxiety and confusion. His system of self-referred beliefs becomes unshakable. If he permits himself doubt about it, he risks precipitation into vast confusion. Finding himself without family, job, and other important props to his sense of identity, increasing doubt may assail the recently hospitalized schizophrenic. These further dispose him toward the adoption of a reliable, if invalid, self-system.

While high variability of self-concept does not distinguish a person as schizophrenic (there was considerable overlapping between groups in this study), it does appear that self-concept disorder is an important feature of many nonchronic schizophrenics and may significantly influence symptom development.

Therapy for the schizophrenic might be substantially improved by closer attention to the patient's self-concept. After obtaining information about how the patient sees himself, the therapist, in the role of rational, consistent, and benign parent surrogate, may help the person to develop and test durable realistic views about himself. The therapist must serve as a sounding board for attempts at self-definition and consensual validator of those self-perceptions which conform to present reality or the potential development of the person. Though the destructive effects of early relationships may never be completely nullified, a brake may be put upon the desperate adoption of new symptoms which lead to further alienation from society and resistance to recovery.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Bruner, 1955.

BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *Journal of Abnormal and Social Psychology,* 1952, **47**, 597–606.

COMBS, A. W., & SNYGG, D. *Individual behavior: A perceptual approach to behavior.* New York: Harper, 1959.

DILLER, L. Conscious and unconscious self-attitudes after success and failure. *Journal of Personality,* 1954, **23**, 1–12.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Reinhart, 1950.

ENGLE, M. The stability of the self-concept in adolescence. *Journal of Abnormal and Social Psychology,* 1959, **58**, 211–215.

KATZ, I. Emotional expression in failure: A new hypothesis. *Journal of Abnormal and Social Psychology,* 1950, **45**, 329–349.

LECKY, P. *Self-consistency: A theory of personality.* Garden City, N. Y.: Island Press, 1945.

ROGERS, C. R. *Client-centered therapy; its current practice, implications, and theory.* Boston: Houghton-Mifflin, 1951.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 1935, **18**, 643–661.

SULLIVAN, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1953.

(Received June 28, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 280–286

INDIVIDUAL DIFFERENCES IN MENTAL ACTIVITY AT SLEEP ONSET[1]

DAVID FOULKES, PAUL S. SPEAR, AND JOHN D. SYMONDS

University of Wyoming

Reports of sleep-onset and nocturnal (EEG Stage 1, rapid eye-movement [REM] period) mentation, CPIs, and thematic fantasy responses were collected from 32 young-adult Ss. In confirmation of recent findings by Foulkes and Vogel, dreamlike mental activity was found to occur with fairly substantial frequency at sleep onset, albeit with wide individual differences. Patterning of personality correlates with hypnagogic and nocturnal dream recall suggests that the former varies directly with waking ego-strength and adaptive flexibility, while the latter varies directly with indications of waking maladaptive symptomatology. A psychodynamic interpretation of these results is offered.

Considerable evidence (e.g., Dement, 1965) has accumulated in recent years relating nocturnal dreaming to a particular phase of EEG sleep, ascending EEG Stage 1 accompanied by intermittent rapid eye movements (REMs). Periods of this stage of sleep recur in a cyclic fashion throughout the night. They are not normally observed at the onset of sleep, however, the first such period generally occurring only after an hour or more of non-REM sleep.

Foulkes and Vogel (1965) have recently demonstrated the occurrence of dreamlike mentation at sleep onset in the absence of an ascending Stage 1 EEG or REMs. Although the content of such mentation may ultimately prove discriminable from content elicited during REM sleep, leaving undisturbed the association of REM sleep with a unique class of mental fantasy, these authors state that, by superficial appearance, it is not always easy to discriminate the sleep-onset dream from the REM-sleep dream.

One of the most striking findings of the Foulkes and Vogel study was the extent of individual differences in the reporting of dreaming at sleep onset and in the association of dream reports with particular phases of EEG and EOG activity. This inter- and intrasubject variability in the reporting of dreamlike fantasy at sleep onset contrasts with the practically universal high frequency of dream-reporting upon awakenings from nocturnal periods of REM sleep. These authors also noted their "clinical impressions" that the subjects who frequently reported dreamlike mentation at sleep onset were less anxious and constricted than the subjects characterized by better controlled and less hallucinatory mental content during this period.

The present investigation was undertaken (*a*) to attempt to confirm, with a much larger subject sample than that of the Foulkes and Vogel study, the frequent occurrence of dreamlike fantasy at sleep onset; (*b*) to determine whether personality variables associated with the hypnagogic occurrence of dreamlike fantasy are in accord with Foulkes and Vogel's speculations on personality differences between sleep-onset dreamers and sleep-onset nondreamers; and (*c*) to examine personality variables differentially associated with hypnagogic and nocturnal (REM) dreaming, in an attempt to assess the possibly different psychodynamic functions each kind of dream might serve.

METHOD

Thirty-two subjects, 16 young-adult males and 16 young-adult females, were recruited from a predominantly freshman course in introductory psychology, receiving class-credit points for their service. Each subject spent 1 night in the laboratory, during which continuous EEG and EOG recordings were made according to the method of Dement and Kleitman (1957). EOG recordings were taken in monopolar fashion from electrodes placed at the outer canthus of each eye.

---

[1] Support for the research herein reported has been received from the Department of Psychology, University of Wyoming, and from the National Science Foundation under Grant GS-860 to David Foulkes. The authors also wish to acknowledge the substantial and generous assistance of Allan Rechtschaffen in the establishment of a laboratory for sleep research at the University of Wyoming, and the helpful comments of Gerald Vogel upon an earlier draft of this report.

Experimental awakenings were made both during the sleep-onset period and during REM sleep. A total of seven awakenings was scheduled for each night in the following order:

Hypnagogic awakenings

1. Sleep-onset descending Stage 1 EEG[2]
2. Sleep-onset alpha EEG with slow eye movements (SEMs)
3. Sleep-onset alpha EEG with SEMs
4. Sleep-onset descending Stage 1 EEG

Nocturnal awakenings

5. REM sleep, 10 minutes after onset of second REM period of the night
6. REM sleep, 10 minutes after onset of third REM period of the night
7. REM sleep, 10 minutes after onset of fourth REM period of the night

The sleep-onset awakenings were made in categories found by Foulkes and Vogel to produce an intermediate level of hypnagogic recall of dreamlike fantasy, hence, in categories especially favorable for a study of individual differences in such recall. Not all subjects completed the entire schedule because of their failure to complete four full cycles of sleep following the initial series of sleep-onset awakenings, but a total number of 78 REM-sleep awakenings were made with 31 subjects as well as the entire number of 128 scheduled sleep-onset awakenings.

Awakenings were made by the experimenter calling the subject by name. When alert, the subject was asked what, if anything, had been going through his mind immediately before the experimenter's call. Upon completion of a spontaneous report, subjects were then asked if the content they experienced was visual ("Was there any visual imagery? What did you see?") and hallucinatory ("Were you, during the experience you've reported, aware that you were observing the contents of your own mind, or did you feel that you were participating in or observing events out in the real world?") and were also asked to classify their state of wakefulness or sleep immediately prior to the experimenter's call ("Were you awake and alert, awake but drowsy, drifting off to sleep, in light sleep, or in deep sleep?"). If the subject had no mental content to report, he was asked to "think on it" for a moment or so, and, if his answer was still negative, he classified his state of wakefulness or sleep immediately prior to the awakening and was then allowed to attempt to return to sleep. All interviews were tape-recorded for later transcription.

The transcriptions of the sleep-onset and REM-sleep interviews were transferred to cards, one awakening protocol to each card. These cards, arranged in a random order, were given to two judges who knew neither the subjects nor the awakening categories associated with the reports contained therein. The judges were to rate the reports on 7-point scales for the appearance of Aggressive and Sexual content and for Hedonic Tone (1-very pleasant to 7-very unpleasant). Judges also coded the subjects' responses indicating their Depth of Sleep (0, awake and alert; 1, awake but drowsy; etc.) prior to awakening, and, employing the subjects' responses to the questions on the visual and hallucinatory aspects of their experiences, rated all reports on the following scale of Dreamlike Fantasy:

0. No content reported, feels mind was blank
1. No content reported, feels something was going through his mind, but forgets what
2. Conceptual content, everydayish
3. Conceptual content, bizarre or unusual topics
4. Perceptual content, nonhallucinatory, everydayish, undramatic
5. Perceptual content, nonhallucinatory, bizarre or unusual, dramatic
6. Perceptual content, hallucinatory, everydayish, undramatic
7. Perceptual content, hallucinatory, bizarre or unusual, dramatic

Reliabilities of the judges' ratings were as follows (averaged subject-ratings for all reports):

	Hypnagogic reports	REM reports
Aggression	.81	.73
Sex	.91	.46
Hedonic Tone	.72	.45
Depth of Sleep	.996	.999
Dreamlike Fantasy	.96	.99

The relatively low reliability for Sex for REM reports reflects the almost total absence of this category of content in our sample of REM reports, while the relatively low REM reliability for Hedonic Tone must reflect the relatively greater ambiguity of hedonic tone in REM content than in hypnagogic content or in wakeful fantasy, since the same judges yielded considerably higher reliabilities in their ratings of hedonic tone for these latter kinds of content than they did for REM content.

Subjects contributed three kinds of information in addition to their reports of hypnagogic and nocturnal mentation. Prior to their night of experimental service, they estimated their usual frequency of dream recall on Antrobus' (1962) scale. "Recaller" scores were computed by converting the subject's response (e.g., a check at "once every night") into a weekly frequency (e.g., 7.00). Subsequent to their night of experimental service, all subjects completed the California Psychological Inventory (CPI) and wrote responses to a "Picture-Story Test" which consisted of cards 3BM and 11 of the Thematic Apperception Test (TAT). CPIs were scored for the standard

---

[2] Descending Stage 1 EEG differs from ascending Stage 1 EEG in several respects: (*a*) it follows wakefulness rather than non-REM sleep; (*b*) it is unaccompanied by REMs; and (*c*) it has not hitherto been thought to be associated with the occurrence of dreaming.

TABLE 1

Mean Ratings of Hypnagogic and Nocturnal Mentation

	Hypnagogic reports			Alpha SEM versus Stage 1 p	REM reports	REM versus hypnagogic p
	Alpha SEM	Stage 1	Total			
Aggression (0–6)	.14	.16	.15	—	.21	—
Sex (0–6)	.14	.05	.09	—	.07	—
Hedonic tone (1–7)	4.00	4.21	4.11	—	4.06	—
Depth of sleep (0–4)	1.53	2.20	1.86	.001	3.64	.00001
Dreamlike fantasy (0–7)	4.46	4.41	4.44	—	4.87	—
Word count (mean number)	74.1	80.1	77.1	—	90.7	—

scales, while TAT responses were rated by the two judges for Aggression, Sex, Hedonic Tone, and Imaginativeness. Reliabilities of these ratings (summed subject-ratings for the two stories) were: Aggression, .92; Sex, .70; Hedonic Tone, .85; and Imagination, .91.

Results

Incidence and Quality of Mental Content Reported at Sleep Onset

Table 1 presents the means of subject means for each of the rating variables applied to alpha-SEM and Stage 1 hypnagogic reports and to REM-sleep reports and for substantive word counts made of each class of report. All intrahypnagogic and hypnagogic-REM mean differences were tested with the Wilcoxen matched-pairs signed-ranks test (two-tail). From the data in Table 1 it is apparent that there is very little manifest instinctual (aggressive, sexual) content in either hypnagogic or REM reports, that mean hedonic tone does not vary significantly between reports of hypnagogic mentation and reports of nocturnal mentation, and that neither quantity (Word Count) nor quality (Dreamlike Fantasy) dimensions successfully discriminate hypnagogic from REM mentation. The only rating dimension which does discriminate descending Stage 1 reports from alpha-SEM reports, and REM reports from hypnagogic reports, is subjectively estimated Depth of Sleep before "awakening," and this is essentially a validation of the experimenter's choice of points at which the subject was awakened.

A more refined analysis, however, does reveal some differences between hypnagogic and REM-sleep reports. Considering reports which meet a relatively strict criterion of dreamlikeness (rated either 6 or 7 by both raters on the scale of Dreamlike Fantasy), a difference emerges between the two classes of awakening. Of 104 such "dreams," 54 were from REM sleep (comprising 69.2% of all REM awakenings) and 50 (29 from descending Stage 1 and 21 from alpha-SEM) from hypnagogic awakenings (comprising 39.1% of all hypnagogic awakenings). This difference was significant (by the Wilcoxen test) at the .01 level. When consideration is further restricted to reports rated 7 by both raters on the scale of Dreamlike Fantasy (i.e., reports unequivocally perceptual-hallucinatory and also either bizarre or highly dramatic), 21.8% of REM reports are so classified, compared to 12.5% of descending Stage 1 reports and 12.5% of alpha-SEM reports. This difference was significant (again by the Wilcoxen test) at the .05 level. "Dreams," in terms of either of two relatively strict sets of criteria, thus occur significantly more often during REM sleep than during the sleep-onset period. They are by no means absent, however, at sleep onset, occurring more than half as often as during REM sleep. In this respect, the data of the present study confirm Foulkes and Vogel's observations of the occurrence of dreaming at sleep onset.

Further analysis also reveals why mean scores on Dreamlike Fantasy do not discriminate REM reports from hypnagogic reports while the frequency of 6 or 7 ratings does so discriminate. REM ratings on this scale tend to constitute a bimodal distribution, with scores clustering either at the high end of the scale, that is, strongly dreamlike recall (69.2% of the judgments of both raters are 6 or 7), or at the low end of the scale, that is, no recall at all (17.9% of the reports are judged by both raters as 0 or 1). While hypnagogic reports are not so often rated

6 or 7 as are REM reports, neither, on the other hand, are they so often rated 0 or 1. Only 3.1% of alpha-SEM reports failed to produce any recall, while the comparable figure for descending Stage 1 was 7.8%. The difference between hypnagogic and REM awakenings in producing no-content reports was significant, by the Wilcoxen test, at the .01 level. Thus, REM awakenings produce both significantly more "dream" reports and significantly more no-content reports than do hypnagogic awakenings. Either recall or the content to be recalled appears to be more of an all-or-nothing phenomenon during REM sleep, while during the sleep-onset period either recall or content is more of a graded process.

The essentially rectangular distribution of sleep-onset ratings on the Dreamlike Fantasy Scale confirms Foulkes and Vogel's observation of substantial individual differences in the occurrence of dreamlike fantasy at sleep onset. Considering reports rated 6 or 7 by both raters on this scale, for example, 8 subjects had a 0% frequency of such reports, 8 subjects had 25% frequency, 8 subjects had 50% frequency, 6 subjects had 75% frequency, and 2 subjects had 100% frequency. Such a distribution, of course, is favorable to the investigation of personality factors associated with individual differences in the occurrence of dreamlike fantasy at sleep onset.

Personality Variables Associated with Sleep-Onset Dreaming

Significant positive Pearson product-moment correlations were observed between the CPI Social Presence and Self-Acceptance Scales and hypnagogic Dreamlike Fantasy ratings, averaged over the two raters and summed over all four hypnagogic reports (.35 and .37, respectively, $p = .05$). A significant negative correlation was observed between such ratings of hypnagogic Dreamlike Fantasy and the CPI Socialization Scale ($r = -.57$, $p = .001$). Hypnagogic dreamers thus seem to have greater social poise, to be more self-accepting, and to be less rigidly conforming to social standards than do hypnagogic nondreamers.

Although ratings of sleep-onset mentation on the Dreamlike Fantasy scale correlated significantly with REM Word Counts ($r = .47$, $p = .01$), they did not correlate significantly with ratings of nocturnal mentation on the Dreamlike Fantasy scale ($r = .20$) nor with other qualitative features (aggression, sex, hedonic tone) of reported REM mentation. It does not appear, then, that hypnagogic and nocturnal Dreamlike Fantasy are highly intercorrelated with one another.

The picture is quite different, however, when hypnagogic and waking fantasy are compared with one another. Hypnagogic Dreamlike Fantasy correlated positively and significantly with all five scores available for Thematic Apperception Test content: with TAT Word Counts, $r = .42$ ($p = .05$); with TAT Aggression, $r = .51$ ($p = .01$); with TAT Sex, $r = .39$ ($p = .05$); with TAT Hedonic Tone, $r = .44$ ($p = .05$); and with TAT Imagination, $r = .40$ ($p = .05$). Hypnagogic mentation and controlled waking fantasy production are, evidently, strongly intercorrelated with one another. By inference, one might conclude that hypnagogic fantasizing is under the same kinds of ego-controls as predominate in the thematic fantasy test, but that neither share as much in common with true nocturnal fantasy as they do with one another. And, in fact, TAT imagination correlates only .22, a value not significantly different from zero, with REM-sleep ratings of Dreamlike Fantasy. Thematic Apperception Test Aggression and Hedonic Tone did correlate significantly, however, with REM-sleep Dreamlike Fantasy ($r = .45$ and $r = .42$, respectively; $p = .01$ and $p = .05$, respectively).

Also supporting the notion that waking ego-functions are more strongly related to hypnagogic fantasizing than to nocturnal fantasizing is the evidence with respect to the CPI Psychological Mindedness Scale. This scale correlated positively ($r = .17$), but insignificantly, with hypnagogic Dreamlike Fantasy, while it was the only one of 18 CPI scales to correlate significantly with nocturnal Dreamlike Fantasy, the direction of the relation being negative ($r = -.40$, $p = .05$) in this case.

The CPI Socialization Scale which correlated so significantly with hypnagogic Dreamlike Fantasy correlated .01 with nocturnal Dreamlike Fantasy. The CPI Social Presence and Self-Acceptance Scales, which also related significantly with hypnagogic

Dreamlike Fantasy, correlated −.16 and .10, respectively, with nocturnal Dreamlike Fantasy, neither of these values being significantly different from zero.

Recaller scores, preexperimental impressions of frequency of dream occurrence, correlated significantly ($r = .35$, $p = .05$) with ratings of nocturnal Dreamlike Fantasy, but .00 with ratings of hypnagogic Dreamlike Fantasy. Evidently, of the two varieties of dream, it is the nocturnal REM-period dream which provides whatever accurate basis there is for these everyday impressions of dream occurrence. The low magnitude of the significant correlation indicates, however, that factors other than variations in the quality of nocturnal fantasy determine everyday impressions of dream occurrence. The significant correlation of Recaller scores with certain CPI scales (Socialization, −.37 [$p = .05$]; Self-Control, −.45 [$p = .01$]; and Good Impression [a fake-good scale comparable to L or K on the MMPI], −.36 [$p = .05$]) which were not significantly associated with nocturnal Dreamlike Fantasy suggests that, apart from any differences in the actual quality of reportable REM-sleep mental content, it is the highly socialized, self-controlled, and defensive person who is least likely to be aware of the nature and extent of such content. Psychological Mindedness, which correlated negatively with the Dreamlike Fantasy of experimentally elicited REM reports, also correlated negatively ($r = -.55$, $p = .01$) with everyday impressions of dream recall.

Differential Personality Variables Associated with Sleep-Onset and Nocturnal Dreaming

From the CPI and TAT variables associated with hypnagogic dreaming and nocturnal dreaming, it appears that the former is more of an ego-function and more directly related to waking ego-strength and flexibility than is the latter.

But perhaps the best method for illuminating personality characteristics differentially associated with hypnagogic and nocturnal (REM) dreaming is the examination and comparison of individual personality test items which best discriminate hypnagogic "dreamers" from hypnagogic "nondreamers" and nocturnal "dreamers" from nocturnal "nondreamers." To this end, an analysis was conducted to determine the CPI items most successful in discriminating responses of subjects rated high in hypnagogic Dreamlike Fantasy from responses of subjects rated low in hypnagogic Dreamlike Fantasy and in discriminating responses of subjects rated high in nocturnal Dreamlike Fantasy from responses of subjects rated low in nocturnal Dreamlike Fantasy. The groups selected for this analysis comprised the top and bottom quarter of the distributions of ratings of Dreamlike Fantasy in each period of sleep. There was little overlap between the REM-sleep and the hypnagogic comparison groups: of the 8 subjects high in REM fantasy, 2 were in the corresponding group for hypnagogic fantasy, 3 in the low group for hypnagogic fantasy, and 3 in neither of the hypnagogic comparison groups; of the 8 subjects low in REM fantasy, 2 were in the corresponding group for hypnagogic fantasy, 2 in the high group for hypnagogic fantasy, and 4 in neither of the hypnagogic comparison groups. The sex-composition of comparison groups was well-balanced for each sleep period (in the high hypnagogic fantasy group there were 5 males and 3 females, in the low hypnagogic fantasy group there were 4 males and 4 females, and both the high and low REM fantasy groups were comprised of 5 females and 3 males).

Table 2 presents the results of the CPI item analyses. Looking first at the items most successful in discriminating high and low hypnagogic "dreamers," it is apparent that several characteristics of the "authoritarian personality" syndrome pervade the entire list (rigidity, conventionality, intolerance, anti-intraception) with the hypnagogic nondreamers' responses closely approximating the standard conception of the authoritarian syndrome (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950).

An inspection of the items most successful in discriminating high and low nocturnal "dreamers" reveals considerably less immediate homogeneity of content than was the case for the items discriminating hypnagogic fantasy groups. It is readily apparent, however, that somewhat different dimensions of personality are operative in the two cases. While the item response "false" characterized the high hypnagogic fantasy group on 7 of the 8

TABLE 2

MOST DISCRIMINATING CPI ITEMS

Endorsement by group high in dreamlike fantasy at sleep onset	Item	Endorsement by group low in dreamlike fantasy at sleep onset
12.5%	112. I set a high standard for myself and I feel others should do the same	87.5%
87.5%	148. I believe women should have as much sexual freedom as men	12.5%
12.5%	367. My home life was always very pleasant	87.5%
0.0%	67. I feel sure that there is only one true religion	62.5%
12.5%	69. I would disapprove of anyone's drinking to the point of intoxication at a party	75.0%
37.5%	131. One of my aims in life is to accomplish something that would make my mother proud of me	100.0%
37.5%	371. I would rather be a steady and dependable worker than a brilliant but unstable one	100.0%
0.0%	404. I am in favor of a very strict enforcement of all laws, no matter what the consequences	62.5%
Endorsement by group high in dreamlike fantasy during REM sleep	Item	Endorsement by group low in dreamlike fantasy during REM sleep
75.0%	105. I am fascinated by fire	0.0%
87.5%	128. It takes a lot of argument to convince most people of the truth	25.0%
62.5%	151. I have strange and peculiar thoughts	0.0%
62.5%	155. A person should adapt his ideas and his behavior to the group that happens to be with him at the time	0.0%
75.0%	300. Police cars should be especially marked so that you can always see them coming	12.5%

most discriminating items for fantasy during that period of sleep, the item response "true" characterized the high fantasy group for all 5 of the most discriminating items for fantasy reported from REM sleep. Perhaps the simplest common denominator of the 5 items to which nocturnal dreamers tended to reply "true" is that all are keyed *false* on several scales of social adjustment or personal growth (105 and 300 on the CPI Responsibility Scale, 128 on the CPI Capacity for Status Scale, 151 on the CPI Self-Control and Tolerance Scales, and 155 on the Psychological Mindedness Scale). These responses of the high fantasizer during REM sleep thus seem indicative of something less than optimal social or personal adjustment. Indeed, in one case ("I have strange and peculiar thoughts"), the manifest content of the discriminating item which is endorsed by the REM-period fantasizer suggests considerable disturbance. Such results and interpretations are consistent with Foulkes and Rechtschaffen's (1964) data demonstrating a pattern of positive correlation between scores on MMPI pathology scales and intensity of recalled REM-sleep mentation. Shifting from the more obvious characteristics of the discriminating items for REM-sleep fantasy to their possible latent meaning, it appears that while the low hypnagogic fantasizer exerts rigidly successful control over his impulse life, the high nocturnal fantasizer shows fascination with impulse life (as symbolized, for example, by fire) in conjunction with weakened ego-control mechanisms (as indicated by the presence of strange and peculiar thoughts and by the concern with evading detection evident in the police-car response and in the strategy of adopting the protective coloring of conformity to the environment in which one finds himself).

DISCUSSION

Dreaming, sometimes difficult to distinguish from REM-sleep dreaming, does occur with fairly substantial frequency during sleep onset. Although, in the present study, "dreams" (i.e., reports which judges agree are of perceptual, hallucinatory, and dramatic content) occurred significantly more often during REM sleep than at sleep onset, they did occur at sleep onset, and more than half as often as they did during REM sleep. As Foulkes and Vogel (1965) have noted, such findings imply that research on the physiological basis of, or on mental activity occurring during, nocturnal periods of REM sleep cannot provide completely satisfactory answers to questions on the physiology and psychology of dreaming.

But there clearly are certain respects in which sleep-onset dream activity differs from nocturnal dream activity. One of these is the susceptibility of the association of dream reporting with EEG/EOG activity to variation from subject to subject. The results of the present study indicated, in confirmation of Foulkes and Vogel's earlier but less systematic observations, that healthy and expansive ego functioning during wakefulness is positively associated with the occurrence of dreamlike fantasy during sleep onset: the hypnagogic dreamers tended to have greater social poise, to be more self-accepting, to be less rigidly conforming to social standards, and to be more adept at producing voluntary waking fantasy than did the hypnagogic nondreamers. These findings may be contrasted with the minimal indications in the present study, and the more substantial evidence collected by Foulkes and Rechtschaffen (1964), that signs of maladaptive or pathological waking response are positively associated with dreamlike qualities of REM-sleep fantasy.

Elsewhere, Vogel et al. (1966) have presented a psychodynamic view of sleep onset. They note that successful sleep onset implies increasing withdrawal of ego cathexes from the external world and propose that this withdrawal has certain unintended consequences, viz., narcissistic, regressive mental content. It is also suggested that individual differences in ego-strength are directly associated with tolerance for such content and, hence, with variations in the dreamlike quality of mentation at sleep onset. The findings of our study are entirely in accord with this view of hypnagogic dreams as ego-processes, initiated and controlled by the self, and as correlates of the adaptiveness of waking ego-functioning.

Vogel et al. also propose that the REM dream is the ego's defensive response to pressures (increased instinctual tensions, etc.) arising outside the ego, rather than a process initiated or directly controlled by the ego itself. The less adequate the ego, the less well-controlled will be the ensuing dream. On this view, it is to be expected that ego-strength will be inversely related with the dreamlike properties of mental experience during REM sleep. The evidence of the Foulkes and Rechtschaffen (1964) study, and to a lesser degree, of the present study, is also supportive of this proposition.

It appears, then, that the hypnagogic dream might profitably be viewed as an ego-controlled excursion into inner thoughts and feelings following the ego's voluntary decathexis of sensory input from the external world. Subjects with rigid defenses against impulse-life tend to resist any encroachment of such regressive mental content upon wakeful levels of ego-functioning, and thus experience little dreamlike content at sleep onset. The nocturnal dream, on the other hand, may be viewed as the ego's involuntary response to unconscious impulses and anxieties which become active during REM sleep. Subjects with egos lacking adequate defenses against impulse-life tend to be overwhelmed by it during REM sleep, hence experience especially vivid REM-sleep dreams.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ANTROBUS, J. S. Patterns of dreaming and dream recall. Unpublished doctoral dissertation, Columbia University, 1962.

DEMENT, W. C. An essay on dreams. In T. M. Newcomb (Ed.), *New directions in psychology II*. New York: Holt, Rinehart & Winston, 1965. Pp. 135–257.

DEMENT, W., & KLEITMAN, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 1957, **9**, 673–690.

FOULKES, D., & RECHTSCHAFFEN, A. Presleep determinants of dream content: Effects of two films. *Perceptual & Motor Skills*, 1964, **19**, 983–1005.

FOULKES, D., & VOGEL, G. Mental activity at sleep onset. *Journal of Abnormal Psychology*, 1965, **70**, 231–243.

VOGEL, G., FOULKES, D., & TROSMAN, H. Ego functions and dreaming during sleep onset. *Archives of General Psychiatry*, 1966, **14**, 238–248.

(Received June 30, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 287–299

SUICIDE AMONG PATIENTS WITH DIAGNOSES OF ANXIETY REACTION OR DEPRESSIVE REACTION IN GENERAL MEDICAL AND SURGICAL HOSPITALS

NORMAN L. FARBEROW
Central Research Unit,
Veterans Administration Center, Los Angeles

AND

THEODORE L. McEVOY
University of California at Los Angeles

Analysis of records of Veterans Administration General Medical and Surgical patients with anxiety or depressive reactions, half of whom had committed suicide, and the rest of whom were matched nonsuicidal controls, showed that suicides and nonsuicides were significantly differentiated on each of 23 feeling and behavior items. Also most of the cases could be categorized in patterns, called (a) Object-Loss, suicides with severe interpersonal conflicts; (b) Involutional, suicides with insult to their psychological integrity; (c) Medical, controls with physical problems; (d) Egoistic, both suicides and controls with few ties to society; or (e) Preegoistic (tentatively identified), younger controls who were possibly in an earlier stage of the Egoistic syndrome. An important measure against suicide is evidence that members of the hospital staff are interested in, and concerned about, the patient.

Depression has been noted as a characteristic symptom in most suicide cases (Farberow, 1950; Masserman, 1947; Mintz, 1961; Shields, 1946; Stengel, 1955). Anxiety has also been indicated as a prominent feature (Farberow, 1950; Farberow & Shneidman, 1960). When these symptoms are so marked that they can be characterized as psychoneurotic syndromes, suicide can be expected to occur proportionally more often among persons so classified.

A survey of cases of suicide[1] occurring among patients in the Veterans Administration General Medical and Surgical Hospitals indicated that a large number of the suicides were patients with diagnoses of anxiety reaction or depressive reaction, both on neuropsychiatric, and on medical and/or surgical wards (Farberow & Shneidman, 1960). This number, however, was smaller than would be expected proportionally from the total number of such cases. A study of patients with a diagnosis of anxiety or depressive reaction who killed themselves was undertaken to determine the identifying personality characteristics and behavioral features of such a group, if any.

PROCEDURES

For both clinical and practical reasons, the two groups of cases, anxiety reaction and depressive reaction, were combined into one group for this study. This was done in order to have a large enough number of cases for statistically meaningful comparisons, and because it was recognized that the two diagnoses are similar enough dynamically often to be combined into one diagnosis of anxiety-depressive reaction.

Preliminary investigation identified 19 anxiety-reaction cases and 24 depressive-reaction cases which had ended with death by suicide. For each of these cases, a control case was selected, matched for age (plus or minus 3 years), sex (all male), race (all Caucasian), religion (Protestant, Catholic, or Jewish), and diagnosis (anxiety reaction or depressive reaction as either primary or secondary diagnosis). The source of data for this study included all the available clinical information contained in the patient's clinical and correspondence folders.

The following steps were used in analyzing the 86 cases (43 suicides and 43 controls): (*a*) A detailed chronological abstract was prepared[2] for each patient from the case history, doctor's and nurse's notes, psychological test results, ancillary therapists' reports, and all other available data. (*b*) A "psychological autopsy"[3] was held for each case by a

---

[1] All cases of suicide among patients in Veterans Administration Hospitals are reposited in the Central Research Unit for the Study of Unpredicted Death, Los Angeles, Norman L. Farberow, Principal Investigator.

[2] The help of Robert Jones, Ronald McDevitt, Dan Sapin, and Bruce Sutkus in preparing abstracts is acknowledged. Calista V. Leonard assisted in the early planning and organization of the study.

[3] The "psychological autopsy" technique has been employed in other investigations by the Central Research Unit (schizophrenia, cancer, and respiratory studies) and has been described in detail elsewhere (Farberow, Shneidman, & Leonard, 1962, 1963).

TABLE 1

CHARACTERISTICS OF ANXIETY AND DEPRESSIVE REACTION PATIENTS

Characteristics	Anxiety		Depressive	
	Suicide $N=19$	Control $N=19$	Suicide $N=24$	Control $N=24$
Age				
Range	30–69	30–70	33–66	28–64
Median	49	48	50	46
Religion				
Protestant	10	10	20	20
Catholic	7	7	4	4
Jewish	2	2	0	0
Length of stay in hospital				
Less than 1 month	7	11	12	11
3 months	8	6	7	7
6 months	2	0	1	3
1 year	2	1	3	2
More than 1 year	0	1	1	1
Diagnosis				
Primary	14	8	15	14
Primary (but listed secondary)	3	3	3	2
Secondary to medical diagnosis	2	8	6	8

group of research psychologists and a psychiatrist.[4] During the "psychological autopsy" the final outcome of the case was withheld from the research staff until after the "clues" had been identified and the predictions about the outcome had been made. (*c*) Summary sheets recorded the findings of the research staff. (*d*) A checklist was evolved grouping the identifying clues into general categories. (*e*) Each of the 86 cases was again reviewed and assessed for the presence or absence of each of these clues. The frequency of the clues among the suicide cases was compared with the frequencies among the control cases (chi-square). (*f*) The cases were further analyzed by patterns which had been tentatively identified among the suicidal and nonsuicidal groups. The clues were also applied to the subgroups to determine their applicability.

The analyses thus provided two kinds of information: (*a*) the identification of clues which (statistically) significantly differentiated anxiety and depressive suicide cases from nonsuicide cases, and (*b*) the identification of patterns which seemed to characterize subgroups of suicidal and nonsuicidal anxiety and depressive cases.

Description of the Population

Groups. The suicidal and the control groups each consisted of 43 patients, 19 with a primary diagnosis of anxiety reaction and 24 with a diagnosis of depressive reaction.[5]

Population characteristics. The population characteristics of the groups may be seen in Table 1.

Age was a controlled variable, so the age distributions for the suicides and controls are similar.[6] Median ages for the anxiety and depressive suicides are 49 and 50, respectively. Thirty-five percent of the suicide cases fall in the decade from 55 to 64, indicating, as in the general population, that suicide is a greater problem within the older age brackets. The religious distribution of the population shows the usual higher number of Protestants, and fewer Catholics and Jews. Table 1 indicates that the suicides and controls were in the hospital for relatively similar periods of time.

The diagnosis of anxiety reaction or depressive reaction was either the primary or the sole diagnosis in 35 of the suicide cases and in 27 of the control cases (a difference not statistically significant). This includes six cases among the suicides and five among the controls in which the diagnosis was listed as secondary, but which, in the judgment of the psychiatrist, was the primary condition. In these cases the listed primary diagnosis was an historical one and did not play a role in the current situation of the patients.

In the remaining cases the diagnosis of anxiety or depressive reaction was secondary to a medical situation. The anxiety suicides had primary medical diagnoses of bronchitis and visceroptosis. The anxiety controls had primary medical diagnoses of ulcer, hemorrhoids, fever, hypochondriasis, stenosis of the larynx, pulmonary fibrosis, and hernia surgery.

The depressive suicides had primary medical diagnoses of asthma, bronchiectasis, hypertension, palsy, cancer, and schizoid condition. The depressive controls had primary medical diagnoses of cardiac hypertension (2), arthritis, diabetes, tuberculosis, and cancer (2).

One hundred twenty-nine anxiety and depressive cases had to be reviewed in order to obtain the required 43 control cases. In only 56 of these 129 cases (43%) was the diagnosis of anxiety reaction or depressive reaction the primary or sole diagnosis. (This contrasts with the ratio of 35 cases among the

---

[4] Anthony J. Brunse served as psychiatric consultant. His help is gratefully acknowledged.

[5] One of these cases committed suicide 65 days after he was discharged from the hospital rolls. To be consistent with the design of the study, however, he is retained in the control population in the analyses.

[6] Comparison of the anxiety suicides and the anxiety controls for age, using the Mann-Whitney U test, yielded a U value of 163, which is not significant at the .05 level of confidence. Comparing the depressive suicides and depressive controls for age, using a Mann-Whitney U test, a z score of .072 was obtained, which is associated with a probability level of .47, not significant. Comparisons were also made between all the anxiety cases versus all the depressive cases. There was no significant difference between the median ages of these two groups.

43 suicides, 81%.) Apparently many patients in a general medical and surgical hospital will receive a diagnosis of anxiety or depressive reaction secondary to a medical problem. Thus, a random sample of "control" patients rather than the matched sample used in this study might have demonstrated many more differences between the suicide and nonsuicide cases. By using a matched sample the task of differentiating between the two groups focused on the more subtle and complex differences rather than the more obvious differences.

Although the groups were not matched on marital status, the differences among them were not significant. Twenty-six of the suicides were married, nine were single, and eight were divorced, separated, or widowed. Among the controls, 24 were married, 6 were single, and 13 were divorced, separated, or widowed.

Table 2 shows that 14 of the patients committed suicide while on leave, pass, or trial visit; 29 killed themselves in the hospital. The choice of suicide method reflected the hospital status of the patient. Most of the patients in the hospital jumped, while most of the patients out of the hospital shot themselves.

RESULTS

Predictions of Suicidal Outcome

The blind predictions of suicidal or nonsuicidal outcome of the 86 cases in staff conferences were analyzed using five raters who had attended and predicted in 35% or more of the case conferences.

The results indicated that the suicides were more readily identified than the controls (79% versus 64%). The researchers were unanimously correct in their predictions in 21 of the anxiety and 18 of the depressive cases. These were about equally divided among suicides and controls in the anxiety group, but the depressive group showed twice as many "correct" predictions for the suicides as for the controls. Six cases were false positives; that is, two of the anxiety cases and four of the depressives, all controls, were incorrectly predicted to be suicides by all members of the research staff. One case did become a suicidal victim 65 days after discharge; another was subsequently hospitalized in a state of deep coma following a suicidal attempt by drug ingestion. Five of these false positive cases had histories of prior suicidal behavior. There were no false negatives, that is, cases actually suicidal which were unanimously predicted to be controls.

TABLE 2

METHODS USED TO COMMIT SUICIDE AMONG ANXIETY AND DEPRESSIVE PATIENTS

Method	In hospital	Out of hospital	Total
Jumping	17	1	18
Cutting	5	0	5
Gunshot wound	0	5	5
Hanging	4	3	7
Drowning	1	2	3
Carbon-monoxide	0	2	2
Jump/train	1	1	2
Strangulation	1	0	1
Total	29	14	43

Clue Analysis

At each psychological autopsy the staff was required to specify the clues that contributed to their decision that the case was a suicide or nonsuicide control. These clues were organized into clinically meaningful groups and compiled into a 43-item checklist (see Table 3). Each of the 86 cases was again reviewed and assessed for the presence or absence of each of these general clues. The suicides and nonsuicide controls were then compared to determine the differences, if any.

Fifteen of the 43 clues significantly differentiated between the suicide and control cases. One additional general clue, "feeling and behavior symptoms prominent during hospitalization," was broken down into 11 clues of specific behavioral symptoms and these were also analyzed. Of these, eight were significantly differentiating.

General Differentiating Clues

The fifteen general clues are as follows:

Poor prognosis. Among the suicide patients, response to therapy was very poor with usually no appreciable alleviation of psychological symptoms. In the cases with prominent medical problems, the long-term prognosis for the disease was always guarded.

Suicidal history. A history of suicidal behavior (i.e., threats, ruminations, attempts) was more common among the patients who killed themselves than among those who did not. Many of those persons had thus "identified" themselves as potentially suicidal patients. Of the controls with suicidal histories,

TABLE 3

CLUES USED FOR AND AGAINST SUICIDE OF ANXIETY AND DEPRESSIVE PATIENTS WITH SIGNIFICANCE OF DIFFERENCE IN OCCURRENCE FOR SUICIDAL AND CONTROL PATIENTS

Clue	Chi-square
Clues for:	
1. Conflict over impending industrial or occupational change	0.19
2. Poor work history	1.28
3. Disturbed by potential marital or "significant other" change	0.56
4. Disturbed sexual adjustment	5.20**
5. Change in family unity	8.12***
6. Discomfort at home or on pass	9.64***
7. Social isolation	4.67**
8. "Character" disorders (alcoholism, drug addiction, etc.)	0.75
9. Disabling organic illness	1.89
10. Deterioration of body image	0.05
11. Poor prognosis	18.90****
12. Felt rejected by hospital	—[a]
13. Negative reaction to hospital staff	3.14*
14. Not gratified in hospital	—
15. Patient "rejected" hospital	6.41**
16. Positive hospital attitude toward patient	2.35
17. Negative hospital attitude toward patient	2.35
18. Severe psychiatric symptoms	7.77***
19. Excessive dependency	0.19
20. Ineffective somatic symptoms	3.44*
21. Ineffective usual defenses	3.26
22. Recent loss of emotional support from significant person or agency	0.23
23. Suicidal history	10.61***
Clues against:	
1. No occupational difficulties	—
2. Family support	—
3. Agency support	—
4. Positive response to agency support	—
5. Social (other) support	0.24
6. No acute psychological crisis apparent	1.69
7. No psychological symptoms	3.26*
8. Recognizes problems as psychogenic	—
9. Has plans for the future	5.65**
10. Adapting well to crisis or illness	5.20**
11. Blames others without insight	1.29
12. History of dissociative reaction	0.14
13. Alcoholism is a refuge	0.62
14. History of "resolved" anxiety or depressive episodes	—
15. Effective obsessive-compulsive defenses	1.02
16. Possible psychotic break	0.43
17. Prominent somatic complaints	0.69
18. Confusion, e.g., toxic, organic, DTs	—
19. Feels well physically	5.20**
20. Lack of suffering from organic causes	0.75
Feeling and behavior symptoms for suicide:	
1. Sleeplessness	20.79***
2. Hostility	3.22*
3. Agitation	12.01***
4. Tension	9.63***
5. Nervousness	13.31****
6. Worries	11.01****
7. Crying, tearful	1.87
8. Marked anxiety	13.31****
9. Marked depression	8.94***
10. Emotional lability	1.08
11. Hopelessness	2.82

[a] No chi-square computed because of insufficient data.
* $p \leq .10$.
** $p \leq .05$.
*** $p \leq .01$.
**** $p \leq .001$.

several might have been considered suicidal clinically even though they had not, at the time of the investigation, killed themselves.

Discomfort at home on pass. Suicidal patients more often found their conflicts exacerbated when released home on leave or pass.

Changes in family unity. Significantly greater among suicides, this clue reflects the lack of support, understanding, or acceptance the patients felt from his family, with no visitors (recorded) while in the hospital, negative changes in marital relationship, etc.

Severe psychiatric symptoms. Although all of the cases, control and suicide, exhibited some psychiatric symptoms (each had a diagnosis of anxiety reaction or depressive reaction), the suicide cases showed significantly more severe symptoms more frequently. These symptoms included marked thinking disorders, such as irrational preoccupations and severe phobias, sometimes even raising the question of incipient psychotic processes.

Negative reaction to hospital. The suicidal patient more often did not experience effective alleviation of his problems or conflicts. Because the hospital did not meet his needs the patient was often demanding, cantankerous, requested AMAs or made attempts to elope and generally was a "problem" patient.

Severe sexual disturbances. These disturbances, more common among suicides, accompanied the indications of increasing marital discord, as expressed in problems of frigidity, impotence, infidelity, etc. The symptoms also reflected a general deterioration in a significant interpersonal relationship.

Social isolation. As a discriminating clue, this was more characteristic of one group of suicides discussed later under the pattern of "egoistic" suicides. Social isolation characterized persons who had no important relationships in their functioning.

Relative ineffectiveness of somatic complaints in binding anxiety or depression. The potentiality for suicide seemed increased when anxious or depressed persons did not possess alternative means of expressing their problem, such as through somatic symptoms.

Ineffectiveness of usual psychological defenses. For the suicides, previously operative defenses, such as compulsive characteristics, denial, projection, etc., no longer seemed effective, and control of anxiety and depression seemed lessened.

Negative reaction to the staff. Not only did the suicides react negatively to their hospitalization (Clue 6) but they were often negative to staff members, and openly hostile, disparaging, and uncooperative significantly more often than control patients.

Adaptation to illness. This clue, contraindicatory of suicide, reflected a successful resolution or adjustment to an illness or psychological conflict. It implies that suicide, as a mode of adjustment, increased in probability as other methods of adjustment or adaptation were not effective.

Physical well-being. This clue, also a contraindicator of suicide, included the notions of general freedom from continuous discomfort and lack of suffering from organic causes. In general, the controls responded well to therapy and hence were not plagued with continuous physical distress.

Plans for future. This clue also contraindicates suicide. The ability to make plans for the future relates in general to evidence of adaptability. It is also related to more favorable prognosis and was significantly more characteristic of the control group.

No psychological symptoms. This "negative" clue characterized the control cases. Freedom from marked or prolonged symptomatology was compatible with rapid and favorable response to treatment.

A general clue, "feeling and behavior symptoms prominent during hospitalization," was examined for its subsymptoms. Eight of these were found to discriminate significantly as clues for suicide between the suicide and the nonsuicide cases. In general, the clues give a picture of highly intensified anxiety and depressive symptoms, such as sleeplessness, agitation, worry, tension, etc.

Drugs

Medications were noted as the cases were reviewed. An attempt to relate the role of the medications to the illnesses and suicide was discarded because of the sparsity of the information. The patients, unlike a group of schizophrenics studied previously (Cohen, Leonard, Farberow, & Shneidman, 1964), were in the hospital a much shorter length of time and data were much less complete. In general, the depressive cases showed little difference in treatment by drugs, with approximately equal numbers of both suicides and controls receiving tranquilizers or hypnotics and sedatives. The anxiety-suicide cases, however, were treated more often with tranquilizers, while the controls tended to receive more hypnotics and sedatives.

Pattern Analysis

Utilizing the prepared abstracts, the data collected during the psychological autopsies, and the original files, all 86 cases were reviewed. Four relatively clear patterns could be established among 65 of the 86 cases, two characterizing primarily suicidal persons, one primarily control, and one found equally often among both suicides and controls. In addition, another group of seven cases seemed to approach a fifth pattern among the suicides, but is only tentatively identified because of its small size. Fourteen residual cases could not be classified in any category. The distribution of the patients in the patterns are shown in Table 4.

The patterns are labeled according to the apparent predominant characteristic of each group. Each of the patterns is discussed in detail below:

TABLE 4

NUMBER OF ANXIETY AND DEPRESSIVE SUICIDES AND CONTROLS IN OBTAINED PATTERNS

Patterns	Anxiety		Depressive		Total		Total
	Suicide	Control	Suicide	Control	Suicide	Control	
Object-loss	7	2	8	1	15	3	18
Involutional	4	1	7	4	11	5	16
Medical	0	9	3	4	3	13	16
Egoistic	6	5	2	2	8	7	15
Preegoistic	1	1	0	5	1	6	7
	18	18	20	16	38	34	72

The Object-Loss Pattern

"I can't give up hope that she will change her mind." The distinguishing characteristic of this primarily suicidal group was that it was interpersonally based with each patient feeling that he was in danger of losing a "love object." It was also true that in some of these cases there was the additional factor that the patient was anxious about his job and felt threatened because of possible occupational failure. However, in no case was there just an industrial conflict, while in all cases there were signs of marital and interpersonal discord. The prominence of vocational conflicts increased roughly with age.

Eighteen cases (15 suicides, 3 controls, $p < .01$) of the total 86 cases in the study fell into this pattern. Most of the patients were younger, with a median age of 36 and range of 30 to 49. Suicidal activity was very prominent in this group, especially during the prehospitalization crisis. Thirteen of the 18 cases (72%) had known suicidal histories.

Fifteen men were married at the time of their hospitalization; the three others were about to be married. The conflicts centered mostly around impending divorce, separation, childbirth rivalry, infidelity of spouse, guilt over extramarital affairs, and "questionable paternity." Along with these marital difficulties was evidence of sexual disturbance such as masturbatory guilt, sexual incompatibility, homosexual panic, feelings of inadequacy and conflict over masculine role, feeling "repelled" by sex, and feeling of "sin." These sexual conflicts were usually acute and were present in the object-loss pattern significantly more often than in the other three patterns ($p < .01$).

The social and marital adjustment was characterized by marked dependency on the spouse or significant other (16 of the 18 cases, $p < .05$). Peer relationships showed feelings of social inadequacy and fears about competing. Work adjustments were typically marred by feelings of guilt, failure, and inadequacy. In every case, the patient was hospitalized because of psychological difficulties (marital and vocational conflicts). In only three of the 18 cases were there also organic or medical problems, and these responded to medical treatment readily.

In the hospital the cases were marked by unrelieved emotional distress, seen in sleeplessness, tension, and agitation. There was unfavorable response to hospital treatment and in many cases their conditions worsened. Although considerably disturbed emotionally, they were, for the most part, free of physical discomfort or pain. Two of the three controls in this group were identified by the staff as false positives. Both had made prior suicidal attempts, and one made a subsequent attempt.

Characteristically, this group of patients experienced trouble on passes (13 out of 18 cases, $p < .05$), frequently returning early or with exacerbated symptoms. This reflected their feeling of loss of interpersonal support and relationships with other significant persons. In addition, these patients were self-centered, obsessively preoccupied with their own problems, unable to entertain any significant change in the status quo, nor to work out a constructive resolution for their future.

Example. A 37-year-old married man was first admitted to a Veterans Administration hospital with the chief complaint of "personality changes in the last four months." According to the patient, he had been constantly employed, carefree, good natured, and happy-go-lucky. However, in the 4 months prior to admission, his personality and behavior had changed and he had become withdrawn, unable to sleep, and had had suicidal thoughts. The patient's wife had been involved in an extramarital affair which the patient had successfully denied to himself, until she finally made it clear to him that she intended to obtain a divorce. The patient had an extremely dependent relationship with his wife which she was apparently unable to tolerate further.

During the hospital stay, the patient was unable to make any plans which would include the loss of his wife. However, she refused all attempts at reconciliation and proceeded with divorce actions. He made a number of veiled suicide threats, such as asking about insurance payments in cases of suicidal death. The patient went home on a week-end pass about 65 days after admission. While on pass, he went to an amusement park and shot himself with a rifle in a nearby deserted shooting gallery.

The Involutional Pattern

This group of patients, principally suicidal cases, was made up of older men who were reacting with strong emotional disturbances to a threat to their psychological integrity. The threat centered primarily around industrial or occupational failures and the fear of losing economic independence. Other "threats" included sudden illnesses or failing health (resulting in severe limitations to occupational or social activities), and recent changes in the marital or family relationship, such as the loss of sex drive, impotence, frigidity in the spouse, or the loss of children through marriage or death. Characteristically, these men had previously achieved a satisfactory social adjustment, had been productive persons, and had only rarely been physically ill. This group was similar to the object-loss cases but without the excessive dependency which characterized the latter group. They had been "successful" husbands and workers for many years in adult life.

Sixteen cases (11 suicides and 5 controls, $p < .10$) fell into the "involutional" pattern. Most of the patients were older (median age 58, range 46 to 66). Nine of the 16 cases (seven suicides and two controls) in this group had histories of prior suicidal behavior.

All 16 patients in this group experienced a severe reaction to changes in their perception of themselves as effective men with ability to be economically independent. Ten suicides and two control patients had been working until just prior to hospitalization. Nine of the suicides were facing imminent retirement, business failure, or loss of work due to "nervousness." In three cases there had been a gradual approach to retirement and, as their effectiveness diminished, there was a corresponding increase in depression and feelings of bitterness and hopelessness.

All the men in this group were married. The conflicts centering in the home situation, along with recent physical illness or psychosomatic problems, contributed considerably to their pervasive feelings of failure and inadequacy. Among the suicides, there were two cases of impotence, three cases of rejection by their wives, two cases in which the child married and left, and three cases with recent illness of the wife.

In the hospital the most prominent syndrome of this group might be best described as an agitated depression. Insomnia, tension, hyperactivity, chronic anxiety, and worry were prominent features. While often pleasant and cooperative on approach, the patients in this group were generally too wrapped up in their own concerns to socialize easily. Behavioral symptoms characteristic of this group were sleeplessness, agitation, tension, nervousness, marked anxiety, and marked depression. Two of the control cases were identified by the staff as false positives. One of these men was a clear suicidal risk who later voluntarily returned to the hospital because he was fearful of killing himself.

These cases were not chronic patients, many of them never having been hospitalized before. This is compatible with the finding that, as a group, these men had achieved reasonably satisfactory work records and were, in many ways, "solid citizens."

Example. A 50-year-old used-car salesman, married and with two grown children, was hospitalized for a severe depressive reaction. The patient had always been "nervous" but had never been hospitalized or treated for any psychiatric disorder. The patient had been in business for himself but had recently developed the unfounded idea that he was "losing out." He was chronically hypersensitive about his lack of formal education and tended to feel inferior because of this. There had been a recent change in the family structure when his daughter had married and his son had left home for college. Never a person with broad interests or social outlets, he had increasingly drawn in upon himself and his home life, and, at the same time, became less able to make major decisions without his wife's support.

The patient entered the hospital with a variety of vague somatic complaints. There was considerable apathy and melancholy with expressions of feelings of worthlessness, unworthiness, and doubts about his sanity. In the hospital, the patient was cooperative but clearly depressed. He was treated with tranquilizers, hydrotherapy, and corrective and occupational therapy but showed little progress. About 37 days after admission, the wife requested that the patient be given a weekend pass since his daughter and son-in-law were visiting. The patient appeared pleased, but while home, without warning, he took the family car, drove to an isolated spot, and hung himself.

The Medical Pattern

The most outstanding feature of this pattern of primarily control patients was that the focal problem seemed to be a medical one with the psychological reactions as secondary manifestations of the illness. When adjustment problems were noted they were moderate, characterological, heterogeneous, and showed no commonality of pattern. A little more than one fourth of the control group fell into this pattern. One may wonder how many of the suicide cases in this group received their psychiatric diagnosis after the fact. However, from a general survey of admissions to general medical and surgical hospitals of cases with primary or secondary diagnoses of anxiety or depressive reaction, it may be estimated that such persons actually account for 50% or more of such admissions.

Sixteen of the 86 cases in the study (13 controls and 3 suicides) fell into this category. The age range for this group was wide, from 30 years to 67 years, with a median age of 56. Within the pattern, the depression cases tended to be older, with a mean age of 59 as compared with 47 for the anxiety cases. Suicidal activity had occurred in all three of the suicide cases in this group but in none of the nonsuicide cases.

Without exception, these men had successfully adjusted to work and to important interpersonal relationships. Thirteen of the 16 patients in this group were married at the time of their hospitalization, including all three of the suicide cases. Of the three not married, one was an elderly widower who made his home with a daughter, and the other was a happily divorced man with 10 children and many close family ties. In only two cases were there indications of marital conflicts, but in neither of these was the hospitalization precipitated by or related to these conflicts.

The one characteristic that unified this heterogeneous group was that each patient was admitted to the hospital for a legitimate medical or surgical problem requiring hospital care. The anxiety or depressive reactions were clearly secondary manifestations of the physical illnesses and these conditions improved along with their general improvement in health.

Almost without exception the nonsuicide patients responded rapidly to therapy, improved quickly in both physical and psychological areas, and were "good" patients—quiet, cooperative, and responsive. Only the three suicide cases differed to some degree, being older men with more serious and lasting illnesses. The suicides seemed to be direct consequences of attempts to cope with the organic illness. The extent of the accompanying depression and suicidal behavior was not recognized.

Example. A 58-year-old man, living with his second wife, had two grown stepchildren who were married and lived outside his home. The patient was a postal employee with many years of service. His health was generally good until age 51 when he entered the hospital for possible pneumonia. There was no significant organic disease found. His complaints of pain were associated with a mild osteoarthritis of the spine. The patient was anxious on admission but rapidly became asymptomatic.

At age 59, the patient reentered the hospital following a coronary infarction. He was somewhat concerned and mildly depressed about the effect of his illness on his position at the post office. However, the heart condition responded well to treatment and social service resolved his occupational concerns. The patient's symptoms rapidly abated and he was discharged with maximum hospital benefit.

The Egoistic Pattern

"The individual ego asserts itself to excess in the face of the social ego and at its expense. . . . Egoistic suicide is the special type of suicide springing from excessive individualism [Durkheim, 1897]."

The most remarkable aspect of this group of patients, distributed equally among suicides and controls, was the small degree of integration of the patients into society. The men were chronic "loners," or social isolates, with apparently few or no important relationships. Their diagnoses might be questioned, for there were relatively few of the reactive features found among the object-loss, involutional, and medical cases. In some cases, they were marginally psychotic with a strong hypochondriacal overlay. However, they did appear to be chronically anxious or depressed and most displayed character-disorder symptoms. They showed obsessive health concern, bodily preoccupations, and self-pity. Prevention of suicide among these patients would have been difficult. Those who killed themselves did so without an observable psychological crisis or specific precipitating circumstances.

Fifteen of the 86 cases in the study (eight suicides and seven controls) fell into this category. The ages ranged from 42 years to 70 years, with a median age of 56 years.

The social isolation of these men was further reflected in their pre- and in-hospital histories. Occupational adjustment was poor. Only one (a control case) of the 15 men was employed at the time of hospital admission. The other 14 men had work records which, for the most part, were both erratic and unskilled. Some had been unemployed as many as 10–20 years. One man had not worked for 40 years of his adult life.

Important familial or other usual close relationships were lacking. None of the men was living with a spouse at the time of the last hospitalization. Several patients had transient superficial contacts with a sibling or a grown child, but did not live with such a person. Three persons had been public wards for over 10 years. Sexual relationships were infrequent, with some sexual contacts with prostitutes noted among some of the younger men. Seven men were divorced, all of them for many years. Two men who were separated had not obtained divorces apparently for religious reasons, but had lived alone for many years.

The patients came to the hospital for a wide variety of reasons, some somatic and medical, and others having vague organic bases or none at all. The hospital picture was generally one of chronic and obsessive bodily preoccupation among these patients. It may be theoretically meaningful that persons whose adjustment was so marginal socially and who invested so little interest or energy in other persons or work should show excessive self-interest and bodily preoccupation. On the ward, these men showed tendencies to remain alone, aloof, disinterested, uncooperative, and negativistic to the staff. The long range prognosis for improvement or changes was poor in 11 of the 15 cases. Interestingly, not one of the patients perceived his problem as psychogenic.

It was difficult to detect clues that would alert the professional person to the probability of suicide among these men. There was a slight tendency toward greater degrees of emotional disturbance among the suicidal patients. Typical of their behavior were notes indicating that they were uncooperative, fearful of surgery, refused preoperative care, refused treatment at times, made a number of AMA requests, refused social service, and demanded therapy or other help intermittently.

Example. A 67-year-old man, born in Norway, came to the United States as a young man, leaving his family in Norway. In World War I, he served in the army overseas for 6 months. In the 1920s, after three children, the patient was divorced by his wife because of excessive use of alcohol. The patient's social adjustment apparently deteriorated and he entered a Veterans Administration hospital for the first time in 1942, suffering from delirium tremens. From 1942 until 1957, he lived either in a Veterans Administration domiciliary or Veterans Administration hospital with only occasional contacts with his daughter.

The patient had a series of medical hospitalizations for a variety of somatic complaints without any active disease or clear-cut precipitating circumstances for hospitalization. During his final hospital stay, the patient was restless, agitated, demanding, and asocial. He was antagonistic toward the staff and tended to try their patience. One day he was missing from his ward and was later found to have drowned himself in a nearby river.

The Preegoistic Pattern

A small but unique group of seven younger patients had many characteristics very similar to the "egoistic" suicide cases, but only one committed suicide. All but one of them were considered suicidal by the research team. These cases ranged in age from 28 years to 46 years, with a median age of 36 years. In this respect they were much younger than the "egoistic" cases. Five of the seven cases had suicidal histories.

None of the seven patients was living with a spouse at the time of the last hospitalization, although four had been married and were now either divorced or separated. All seven were decidedly passive and dependent in their social adjustment. Only two of the seven men had an adequate work record; the rest were economically dependent on family or pension. Although younger men, they were "beat" and appeared to have little interest in anything. They indulged excessively in alcohol or drugs and were preoccupied with bodily symptoms and complaints.

Residual Cases

Fourteen cases (five suicide and nine control) could not be placed into any of the patterns identified. Among the suicides, either the data were too scarce for classification or the patients tended to display features common to several of the patterns described. The control cases, mostly younger men in their forties, had entered the hospital with a variety of somatic complaints but with no organic illness. Each seemed to look to the hospital for relief of some minor concern or as a refuge from a current crisis. They displayed little symptomatic disturbance in the hospital and recovery was, in most cases, both quick and uneventful. Six of the nine control cases were readily identified as nonsuicidal.

DISCUSSION

Suicide occurs most often in persons who are in a disturbed emotional state. The task of predicting suicide potentially among anxious or depressed neurotic patients, by definition emotionally perturbed, becomes an especially difficult one. Even among such persons, accurate prediction is relatively infrequent. Its infrequency, however, emphasizes its seriousness and widespread impact when it does occur.

The task is one of identifying, within an already more susceptible group, those patients who might have higher suicidal potentiality. One goal of this study was to facilitate this identification by the specification of clues and patterns more frequently associated with suicidal behavior. Such patients might then be marked for special suicide prevention measures. It is probably true that the use of the clues and patterns might tend, in some instances, to encompass too many patients. However, this may be neither inappropriate nor inefficient. The delineation of a group with probably higher suicidal potentiality does not assume that all such patients would kill themselves. Even among patients identified by a prior suicide attempt (Stengel & Cook, 1958) or other prior suicide behavior (Eisenthal, 1965; Fox, 1962) less than 10% go on to commit suicide. However, it is also true that suicidal events are more probable among persons in this group than among other nonsuicidal patients or nonpatients (Farberow & Shneidman, 1960). Once the more susceptible group has been identified, it follows that, as in many similar examples of physical disease, further intensive exploration can be made and appropriate preventive measures can more effectively be applied.

Some remarks about the results may be made:

1. In general, for identification purposes the suicidal patient is the more severely disturbed patient, both psychologically and behaviorally. Although all the patients had psy-

choneurotic diagnoses, the suicide patient had psychiatric symptoms which were more severe, with more depression, anxiety, tension, hostility, and agitation than the controls. Where there were physical complications, such as an illness, infection, or somatic disorder, there seemed to be no relief from treatment, thus adding to the discomfort and distress of the patient. Withdrawal, agitation, sleeplessness, sexual disturbances, etc., exacerbated already deteriorated relationships. Withdrawal from, or insistent, intrusive demands on both staff and other patients reflected and produced strong negative feelings.

How useful is the general clue that the suicidal patients are the more seriously disturbed patients? Many neuropsychiatric patients are severely disturbed at one time or another. However, it is unlikely there are many patients, especially among the neurotics, who show such marked aberrations for a prolonged time. When the other clues are added, such as age, prior suicidal behavior, current self-destructive ideation, feelings of hopelessness even in the hospital, the possibility of identifying a potential suicidal acting-out increases greatly.

2. Another important suicidal indicator was a history of prior suicidal behavior, including thoughts of death, fantasies of rescue, expressions of "having had it," as well as an actual attempt, as noted in previous studies (Farberow et al., 1962, 1963). Apparently, a person who has used this mode of behavior in the past is likely to use it again, and to carry the act through to completion.

3. Most of the suicide cases in this study were patients who were on neuropsychiatric wards within general medical and surgical hospitals. Others, however, entered the hospital for a physical illness and, while in the hospital, developed a neurotic reaction. Thus, eight out of the 43 suicide cases (18%) had a diagnosis of anxiety or depressive reaction as the secondary diagnosis to a physical illness. The physician and other hospital staff on the ward, not just the psychiatrist and neuropsychiatric personnel, must be alert to the danger of suicide among their patients who develop severe anxiety and depression reactions. When this occurs, especially among older patients, a psychiatric consultation would be helpful in evaluating the suicidal danger. Besides giving the patient evidence of interest and concern on the part of the hospital staff, it helps the physician with recommendations and gives him the feeling of sharing the responsibility in the making of what might be a critical decision.

4. More than one third of the suicides were 55 years or older. This is of interest in that, generally, it has been thought that neurotic cases tended to be younger. It thus paralleled the often-reported finding among the general population that the suicidal danger increases with white males as they grow older (Dublin, 1963).

Within the tentatively identified patterns, there are theoretical as well as practical observations. In terms of age, the object-loss group was the only consistently young group (median age of 36). The involutional and egoistic groups were almost all older (median ages of 58 and 56, respectively). While the medical group included some younger people, it was also predominantly older (median age of 56).

One might speculate about a psychodynamic relationship between the object-loss and involutional groups. Despite the age difference there were marked similarities. In both groups there were severe blows to self-esteem and psychological integrity stemming either from the direct interpersonal relationship with a love-object or from functioning in the occupational area. In both groups there were feelings of inadequacy, worthlessness, inability, and uselessness. In the younger group marital and sexual problems predominated, and some, but fewer, occupational problems. In the involutional group occupational difficulties predominated, with some, but fewer, marital and sexual problems.

Patients in the involutional group were not as passive dependent as those in the object-loss group. They seemed to have lived their lives with less disturbance and with obviously greater adjustment up to the time their current difficulty arose. While still essentially passive-dependent persons, they had been able to find appropriate life partners and work, and thus to effect a "good" adjustment for most of their lives. When a severe blow to their self-esteem occurred, either through the loss of a job, demotion, through a physical illness which pushed them into a dependency state, or through some sexual difficulty, for example, loss of virility, their defenses were shaken. They reacted with one of the more marked clues for suicide, an agitated depression which could not be relieved, filled with preoccupation, sleeplessness, tension, and worry.

The egoistic group was a much more puzzling group in that their disturbed psychological states seemed to occur without any significant precipitating event. These were borderline or schizoid persons who seemed more to exist on the fringe of society rather than to live in and be a part of it. In the hospital, too, they remained aloof from the staff and other patients, continuing their noninteractive ways. Within this group, although there were as many suicides as there were controls, there seemed to be a few recognizable differences between them. The suicides made many more demands from the hospital. They were excessively dependent, incapable of being gratified, complaining, manipulating, protesting, and rejecting. The hospital was no haven, but probably no place could have been. The suicidal patients within this group also seemed to become suicidal without precipitating cause. However, within the hospital, the two clues which seemed most useful for identification were the presence of prior or current suicidal behavior, either in threats or ruminations, and the excessive complaining, dissatisfied behavior they exhibited.

The medical group probably makes up around half of all those who receive either a primary or secondary diagnosis of anxiety or depressive reaction in general medical and surgical hospitals. Characteristically their emotional reaction was focused around a physical problem which apparently made them anxious and depressed enough to warrant a neuropsychiatric diagnosis. But they differed from the other groups, especially the involutional, in one important aspect. In relation to their physical illness, they were relieved by the treatment they obtained and their symptoms subsided. The medical group also stood out from the other patients in that their interpersonal relationships were satisfactory, their marital relationships secure, and their sexual relationships gratifying. When emotional strain occurs between partners, a physical illness tends to take on many more subtle dynamic implications and may well become much more recalcitrant to treatment. Within a secure relationship, the physical illness tends to respond to treatment, anxiety diminishes, and relief can occur.

The preegoistic group of patients closely resembles the egoistic patients in life style. These groups differed from each other, however, in two aspects of their psychosocial adjustment: the preegoistic men were much younger and they were still on the fringe of social interaction. Their (lack of) social adjustment had not yet jelled or become brittle or chronic. The preegoistic men were also still close enough to remnants of relationships to resort to communication and "cries for help." Perhaps the more isolated and older "egoistic" men gave such little warning of their intentions because they lacked any important persons with whom to communicate.

Although the results of this study were obtained on a specific group, veterans with diagnoses of anxiety reaction or depressive reaction in general medical and surgical hospitals, the conclusions seem consistent with and generalizable to other public and private hospital populations (Chapman, 1965; Kobler & Stotland, 1964). In general, suicide may be seen as an aberrant, unusual, and complex phenomenon, arising out of intense inter- and intrapersonal needs. A feeling of identity and a meaningful self-concept are most important: When these are threatened, the very foundations of man's reasons for existence seem shaken, his hopes recede, and a sense of emptiness and loss pervades the individual. Tension mounts, straining the usual coping mechanisms and defenses. The struggle within each person becomes a measure of his tension tolerance. Because this level of tolerance fluctuates, the picture will vary and we will see "suicidal" persons swing up and down in their depressive feelings and emotionally disturbed behavior.

While the above formulations would seem to apply primarily to persons in the object-loss and involutional patterns, they also fit in part to those in the egoistic pattern. Although these men have largely withdrawn from society and their source of self-esteem does not depend so much upon relationships with others, they seem highly self-centered with a focus on bodily functioning and personal needs. When disruptions occur in these areas, they too react with self-destructive behavior.

Our results do not tell us why suicide is selected as a way out rather than some other resolution. To understand this would require knowing much more in depth about each individual in the study. Our results do tell us that, for the most part, suicide is preventable. Of most significance is that as the person struggles with his feelings, overt behavioral signs appear. These help identify and delineate the suicidal patient. Second, there is a fluctuation in the self-destructive impulses so that there is usually time and opportunity for intervention to occur. Third, ambivalence is always present and the man communicates his wish to live and to be saved at the same time he declares that he wants to die. Fourth, those measures aimed at restoring and supporting feelings of self-esteem, such as interest, concern, recognition, and acceptance on the part of the staff seem crucial. As Kline states:

> Warm rapport and time spent with the patient, even though it does not seem productive of important "insights," is the most valuable type of psychotherapy. It is not necessary constantly to attempt to stimulate the patient to conversation or participation. Simple human presence is important of itself. It is to be expected that the patient will over and over again ask for reassurance and although this will eventually become boring or even irritating to the personnel it is important that such reassurance be provided [Kline, 1961, p. 19].

Encouragement of the patient into as many other activities and adjunct therapies as possible is important. Direct discussion of suicidal tendencies with the patient helps give him the feeling he is understood.

It is also important that the family be integrated with the treatment process. The physician, psychiatrist, or psychiatric social worker can play a crucial role in informing the family about the patient's condition and the suicidal danger, and in enlisting their aid. If the family cannot help, this must be recognized and the patient made aware of the probability that he cannot find, and should not seek, support there. Discharge to the family, or to the same environment which fostered his disturbance, without any change having occurred in that environment, merely invites a recurrence of the disturbance and potential suicidal acting out.

REFERENCES

CHAPMAN, R. F. Suicide during psychiatric hospitalization. *Bulletin Menninger Clinic,* 1965, **29**(2), 35–44.

COHEN, S., LEONARD, C. V., FARBEROW, N. L., & SHNEIDMAN, E. S. Tranquilizers and suicide in the schizophrenic patient. *Archives of General Psychiatry,* 1964, **11**, 312–321.

DUBLIN, L. I. *Suicide: A sociological and statistical study.* New York: Ronald Press, 1963.

DURKHEIM, E. *Le suicide.* J. A. Spaulding & G. Simpson (Trans.) New York: Free Press of Glencoe, 1951. (Original 1897)

EISENTHAL, S. Follow-up of neuropsychiatric hospital patients placed on suicide observation status. 1965, Unpublished Technical Report, Veterans Administration Central Office.

FARBEROW, N. L. Personality patterns of suicidal mental hospital patients. *Genetic Psychological Monograph,* 1950, **42**, 3–79.

FARBEROW, N. L. Summary. In N. L. Farberow & E. S. Shneidman (Eds.), *The cry for help.* New York: McGraw-Hill, 1961.

FARBEROW, N. L., & SHNEIDMAN, E. S. Interim Unpublished Report No. 3, 1960, Veterans Administration Central Office.

FARBEROW, N. L., SHNEIDMAN, E. S., & LEONARD, C. V. Suicide—evaluation and treatment of suicidal risk among schizophrenic patients in psychiatric hospitals. *Veterans Administration Medical Bulletin,* 1962, No. 8.

FARBEROW, N. L., SHNEIDMAN, E. S., & LEONARD, C. V. Suicide among general medical and surgical hospital patients with malignant neoplasms. *Veterans Administration Medical Bulletin,* 1963, No. 9.

FOX, R. Help for the despairing. *Lancet,* 1962, **2**, 1102–1105.

KLINE, N. S. Comprehensive therapy of depressions. *Journal of Neuropsychiatry, Supplement* No. 1, February 1961.

KOBLER, A. L., & STOTLAND, E. *The end of hope.* New York: Free Press of Glencoe, 1964.

MASSERMAN, J. H. A note on the dynamics of suicide. *Diseases of the Nervous System,* 1947, **8**, 324–325.

MINTZ, R. S. *Detection and management of the suicidal patient.* Chicago: Disease-a-Month Year Book, Medical Publishers, July 1961.

SHIELDS, E. A. Depression, then suicide. *American Journal of Nursing,* 1946, **46**, 677–679.

STENGEL, E. The risk of suicide in states of depression. *Medical Practitioner,* 1955, **234**, 182–184.

STENGEL, E., & COOK, N. *Attempted suicide, its social significance and effects.* London: Chapman Hall, 1958.

(Received July 13, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 300–303

PERSISTENCE OF THE WATERFALL ILLUSION AFTEREFFECT AS A TEST OF BRAIN DAMAGE[1]

AGLAIA EFSTATHIOU AND RICARDO B. MORANT
Massachusetts Institute of Technology *Brandeis University*

Without special conditions and/or instructions to encourage continued fixation on the center of the stimulus object, 19 brain-damaged patients (chronic brain syndrome associated with alcoholism) performed less well on the Spiral Aftereffect Test than did a normal control group. However, they performed as well as normals on the Waterfall Illusion Aftereffect Test which does not require continued fixation in order to elicit apparent after-movement. The results support the interpretation that the reported disturbance in the brain-damaged to perceiving apparent aftermovement is due to their relative inability to hold continued fixation.

Brain-damaged subjects report a Spiral Aftereffect (SAE) under experimental conditions and with instructions designed to encourage constant fixation of the center of the spiral. Previous reports of their inability to experience an SAE can probably be attributed to the inattentiveness of the brain-damaged and not to any specific neurological damage which affects the experience of movement aftereffect (Morant & Efstathiou, 1966). The present paper is concerned with another experience of movement aftereffect in the brain-damaged—the so-called waterfall illusion.

If after looking at the moving water of a waterfall a stationary terrain is then fixated, it is experienced as moving upwards. This experience, termed the "waterfall illusion" (Adams, 1834), can readily be demonstrated in the laboratory with a moving set of parallel lines. The waterfall illusion aftereffect (WIAE) depends less critically on fixation than the SAE (Goldstein, 1958; Spitz, 1958). With the spiral, if the motion is radially expanding, fixation must be maintained on the center in order to obtain the aftereffect of radial contraction. Fixation on the periphery results in a movement aftereffect whose direction is different depending on the part fixated. And if fixation is systematically changed in different parts around the periphery, no aftermotion at all results (Morant & Efstathiou, 1966). The WIAE, however, is unidirectional; if the waterfall movement is down, the apparent motion is up, and if up, the apparent motion is down. These relations hold true irrespective of the part of the waterfall attended to. It follows, therefore, that brain-damaged subjects should perform significantly better on the WIAE than on the SAE test if left to their own devices and not continually reminded to fixate. It follows, also, that if the change of fixation does not cancel out or reduce the WIAE, then the duration of the aftereffect can be related to exposure time without concern to whether or not constant fixation on one point was held.

The accurate measurement of the aftereffect on the waterfall might help clarify the discrepant findings relating duration of movement aftereffect to brain damage. For example, Spivack and Levine (1957) reported that brain-damaged adolescents perceived the SAE longer than normals, and, in a subsequent study (Spivack & Levine, 1959), extended their finding to include brain-damaged patients between the ages of 14 and 45. They suggested that "perhaps the extended duration . . . is based on some analogous process of overcorrection by pathological tissue for whatever process is set up by the spinning spiral [Spivack & Levine, 1959, p. 224]." Gallese (1956) and Holland and Beech (1958), on the other hand, reported that the SAE is shorter in brain-damaged than in normals. To round out the picture, Page, Rakita, Kaplan, and Smith (1957) reported no significant differences in duration between adult brain-damaged subjects and normal controls. These di-

---

[1] This investigation was supported by Research Grant M-3658 from the National Institute of Mental Health, Public Health Service.

verse findings may be due to the lack of fixation controls in the SAE tests. Assuming this is so, then the WIAE test, minimizing as it does the importance of fixation, should prove a more reliable instrument in comparing the duration of aftermovement reported by brain-damaged patients and normals.

METHOD

Apparatus

A standard SAE test apparatus consisting of a black 920-degrees Archimedes spiral was used. The spiral was printed on a white cardboard disk 7¼ inches in diameter (15% black and 85% white) and rotated at 80 rpm. The device for inducing the waterfall illusion was specially made and consisted of an upright aluminum panel with a square aperture. A plastic striped belt tightly stretched over two rotating drums (80 rpm) moved past the aperture in a vertical path with a linear speed of 0.78 inch/second. The moving belt, 7 inches wide, constituted the "field" with an angular speed of 28.1 minutes of arc/second. It was made of black plastic electric masking tape, 0.49 inches wide, pasted horizontally onto a white belt (50% white and 50% black). The spiral and waterfall devices were presented separately to the seated subject at a distance of 8 feet. Each apparatus was supported on an adjustable stand and presented at eye level.

Procedure

The instructions given for the SAE were similar to those used by Price and Deabler (1955). During the exposure period, the subject was asked to describe what he saw. When the rotation was stopped, he was asked again, "Now what do you see?" and his answer recorded. The importance of fixation was not stressed. With the waterfall, the subject was instructed to look at the moving belt and describe what he saw when it stopped. If he reported aftermovement he was asked to indicate its termination. The duration of aftermovement was measured with a stopwatch from the time the subject reported seeing it to the time he said it stopped.

Four trials were run with the spiral: in two, the spiral rotated clockwise and in two, counterclockwise. The order of presentation of direction of rotation was random for each subject. Four trials were also given with the waterfall (downward movement). Both the spiral and the waterfall were viewed binocularly under standard fluorescent room lighting.

Patients were interviewed the day before testing and told that the two tests to be administered were part of an eye examination. Both the spiral and waterfall illusions were shown to the patients and their ability to describe their motion evaluated. The spiral test was administered 1 day, and the waterfall test 3 days later, the spiral always being administered first. (The spiral test was readministered to most of the subjects 3 weeks following termination of the experiment. The scores obtained did not differ from those on first testing.)

TABLE 1

DISTRIBUTION OF SPIRAL AFTEREFFECT SCORES (4 TRIALS)

Subject group	Number of times spiral aftereffect reported				
	0	1	2	3	4
Normals ($N = 19$)	0	0	1	0	18
Brain-damaged with Korsakoff psychosis ($N = 9$)	6	2	0	1	0
Brain-damaged without Korsakoff psychosis ($N = 10$)	6	0	2	1	1

Subjects

Nineteen male patients between the ages of 40 and 57 years from Boston State Hospital composed the experimental group. All had a hospital diagnosis of chronic brain syndrome associated with alcoholism. Of these, nine also suffered from Korsakoff psychosis. Patients were excluded from the study if their medical records (*a*) were incomplete, (*b*) indicated a diagnosis of pathology associated with other than alcohol intoxication or with functional disorders other than Korsakoff psychosis, (*c*) indicated visual defects of any sort. All patients who met the above criteria and who were cleared by the neurologist in charge as sufficiently cooperative to undertake the experiment were studied. Nineteen hospital employees and university graduate students, ranging in age from 20 to 52 years, served as the control group.

RESULTS AND DISCUSSION

Table 1 presents the distribution of scores obtained by the normals and the brain-damaged patients on the SAE. It is evident that without special conditions and/or instructions to encourage continued fixation, a large number of the brain-damaged patients failed to report an SAE. All subjects, both brain-damaged and normals, reported a WIAE on every trial. The mean length of time the WIAE was experienced was for the normals, 16.1 seconds; for the brain-damaged without Korsakoff psychosis, 14.0 seconds; for the brain-damaged with Korsakoff psychosis, 14.3 seconds. Although the average aftereffect experienced by the normal subjects is longer than that of the brain-damaged, the difference is not statistically significant (chi-square not significant at the .05 level of confidence).

These results are congruent with those reported by Morant and Efstathiou (1966). From both studies it may be concluded that, at least in the brain-damaged cases studied: (*a*) If proper measures are taken to assure fixation on the SAE test, the brain-damaged function at about normal level. If such measures are not taken (as in the present experiment), the brain-damaged will not experience SAE movement as frequently as normals. (*b*) If a movement inducing device is used (the waterfall) which does not require fixation to induce an aftermovement, the brain-damaged function at about normal level. (*c*) Duration of aftermotion is no better an indicator of specific brain damage than is frequency of report.

Teuber and Bender (1949) reported that objects shown to patients with acute penetrating injuries to the occipital and parieto-occipital lobes are perceptually displaced from the part of the visual field corresponding to the injured area. They related this finding to the short-term displacements Köhler and Wallach (1944) reported in their studies on satiation and consequent figural aftereffects. Others, however, have tried to draw broader relations between figural aftereffect theory and brain damage. Shapiro (1953, 1954) theorizes that brain damage has the specific effect of permanently exaggerating the inhibitory effects of the remaining brain tissue. Eysenck (1955), drawing heavily on Hull's concept of reactive inhibition and to some extent on Pavlov's work, considers satiation analogous to reactive inhibition. Accordingly, he suggests that the extent of "cortical inhibition" after afferent stimulation is related to the position of the subject on the extraversion-introversion scale. High inhibition and short-lasting perceptual aftereffects are predicted in hysterics, extraverts, and the brain-damaged. Klein and Krech (1952) also draw a parallel between brain damage and oversatiation but they predicted that the decreased conductivity of the brain-damaged should lead to longer lasting perceptual aftereffects. These studies used either the palpating of a tapered wooden block ("kinesthetic after effect") or the viewing of a rotating spiral as the stimulus to induce an aftereffect. Unfortunately, the kinesthetic aftereffect is a particularly poor choice unless used with proper controls (Broadbent, 1958). As has been noted, the same is true of the SAE since the adequate control of fixation is critical. With the WIAE, where fixation is not critical, the brain-damaged function at about normal level.

It may well be that the perceptual impairments attributed to the brain-damaged of the type referred to above are comparable to the figural, kinesthetic, and movement aftereffects of the type reported by Köhler and Wallach (1944) and found in normals. Better controlled aftereffects stimuli are required, however, if the relationship is to be evaluated fairly.

REFERENCES

ADAMS, R. An account of a peculiar optical phenomenon seen after having looked at a moving body, etc. *London and Edinburgh Philosophical Magazine and Journal of Science,* 1834, **5**, 373–374.

BROADBENT, D. E. Extraversion and the kinaesthetic figural after-effect. *Bulletin of the British Psychological Society,* 1958, No. 35. (Abstract)

EYSENCK, H. J. Cortical inhibition, figural after-effect and the theory of personality. *Journal of Abnormal and Social Psychology,* 1955, **51**, 94–106.

GALLESE, A. J., JR. Spiral aftereffect as a test of organic brain damage. *Journal of Clinical Psychology,* 1956, **12**, 254–258.

GOLDSTEIN, A. G. On the after-effects of the "waterfall" and "spiral" illusions. *American Journal of Psychology,* 1958, **71**, 608–609.

HOLLAND, H. C., & BEECH, H. R. The spiral aftereffect as a test of brain damage. *Journal of Mental Science,* 1958, **104**, 466–471.

KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proceedings of the American Philosophcial Society,* 1944, **88**, 269–357.

KLEIN, G. S., & KRETCH, D. Cortical conductivity in the brain-injured. *Journal of Personality,* 1952, **21**, 118–148.

MORANT, R. B., & EFSTATHIOU, A. The Archimedes spiral and the diagnosis of brain damage. *Perceptual & Motor Skills,* 1966, **22**, 319–397.

PAGE, H. A., RAKITA, G., KAPLAN, H. K., & SMITH, N. B. Another application of the spiral aftereffect in the determination of brain damage. *Journal of Consulting Psychology,* 1957, **21**, 89–91.

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *Journal of Consulting Psychology,* 1955, **19**, 299–302.

SHAPIRO, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *Journal of Mental Science,* 1953, **99**, 394–409.

SHAPIRO, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of "apparent motion." *British Journal of Psychology,* 1954, **45**, 58–67.

SPITZ, H. H. Neural stimulation in the spiral aftereffect and similar movement aftereffects. *Perceptual & Motor Skills,* 1958, **8**, 207–213.

SPIVACK, G., & LEVINE, M. The spiral aftereffect and reversible figures as measures of brain damage and memory. *Journal of Personality,* 1957, **25**, 767–778.

SPIVACK, G., & LEVINE, M. Spiral aftereffect and measures of satiation in brain-injured and normal subjects. *Journal of Personality,* 1959, **27**, 211–227.

TEUBER, H. L., & BENDER, M. B. Alterations in pattern vision following trauma of occipital lobes in man. *Journal of General Psychology,* 1949, **40**, 37–57.

(Received July 19, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 304–309

AN ACTION TEST OF HYPOTHESES CONCERNING THE ANAL PERSONALITY[1]

GEORGE C. ROSENWALD
University of Michigan
GERALD A. MENDELSOHN
University of California, Berkeley
ALAN FONTANA AND ALEXIS T. PORTZ
University of Michigan

Psychoanalytic characterology posits that distinctive traits which are associated with conflicts in various impulse areas will be manifest in extreme degrees when the conflict is inadequately resolved. 48 Ss performed a hand-eye coordination task (action test) in an odorous dirty medium reminiscent of fecal matter. Poor performance of this task was taken as an index of ineffective defenses against anal impulses. It was predicted that indecisiveness and narrow intellectualism would be associated with ineffective defenses. Ss with unsuccessful action tests showed slower decision making and a selective sensitivity for anal concepts. Questionnaire study indicated that they also profess greater discomfort in areas which are characteristically troublesome for anal personalities.

Psychoanalytic study of individuals with prominent fixations on the anal stage of psychosexual development has yielded the impression that their characteristic defense tactics frequently result in an intellectualized approach to life problems, emotional issues, and interpersonal relationships. It seems, furthermore, that while the mere presence of anal fixations may lead to a sound, creative, and subtle *intellectual mentality,* the latter may decompensate into an *intellectual disturbance* marked by emptiness, ritualism, and pedantry, when defensive efforts are weakened or impulses gain in strength. This has been observed in adolescents in particular (A. Freud, 1936). A second normal compulsive trait often discerned in anally fixated people is that of cool and impassive *deliberateness* in taking action and in making decisions. Its pathological extreme are the paralyzing doubt and *indecisiveness* resulting from decompensated or overly rigid defenses against anal impulses. Normal intellectuality and deliberateness are usually ego-syntonic and adaptive personality traits, whereas their pathological extremes are often experienced as symptoms requiring amelioration. The present study reports an action test of the prediction that relatively ineffective defenses against anal impulses are associated with indecisiveness and intellectual disturbance.

Previous research dealing with anal personality traits has, on the whole, been discouraging. Although one or two factor-analytic investigations have isolated *traits* corresponding closely to the informal impressions reported in the clinical literature, it has not been possible to assign individuals to an anal *type* in accordance with casual clinical usage.

Barnes (1952) isolated two factors, one of orderliness, reliability, and law abidance, the other of sadism, and this accords well with the psychoanalytic formulation. Earlier still, Sears (1936), in his study of trait attribution, discovered that the traits of stinginess, obstinacy, and disorderliness tended to be correlated in the evaluation of subjects by their peers. This too is theoretically consistent.

Schlesinger recently (1963) collated anality questionnaires employed in various experimental investigations, including that reported in the present study, and submitted a total of 154 items to factor analysis. Among the first 12 factors which she extracted were regularity and meticulousness, retentiveness as a style of

---

[1] This investigation has been aided by a grant from the Foundations' Fund for Research in Psychiatry.

life, obstinacy, frugality, concern about dirt and contamination, anxiety over possible loss of control, and sensitivity to smells. Once again, the intercorrelations of factor loadings tend to confirm clinical and theoretical preconceptions, and the extracted factors led to telling results in an investigation of personality and vocational choice.

Although these findings suggest that a single questionnaire could be constructed for the assessment of anal personality characteristics, this would not serve the purposes of the present study, since the qualitative analysis of defensive operations is the chief variable under consideration.

The action test utilized in the present study consists of a fecal-like stimulus and yields a measure of the effectiveness with which the subject coped with the stimulus. This performance criterion was employed to measure defense effectiveness because, according to psychoanalytic theory, effective defenses, in addition to reducing anxiety, also promote adaptation, whereas ineffective defenses do neither. Accordingly, relative task success was taken to reflect an adaptive strength, and relative failure, an adaptive weakness. Another part of the study made use of a questionnaire assessing anxiety about anal stimuli. In this way, defense effectiveness could be approached from both sides. More specifically, the action test does not attempt to determine whether the subject is anally fixated or even whether he is in conflict about anal impulses. Rather, it seeks to evaluate the adequacy with which he resolves whatever anal conflicts are present. Thus, it makes no distinction between individuals without noteworthy anal fixation and individuals who are flexibly defended against anal impulses. The questionnaire already mentioned does, however, make this distinction by means of items which tap stable anal character traits. Finally, it should be noted that rather different personality pictures result from the *choice* as well as the *effectiveness* of various defense mechanisms. However, no effort is made in the present study to compare, say, repression with reaction formation or projection. Only the degree of defense effectiveness, in terms of performance success, is measured.

METHOD

The subjects were 48 male college freshmen and sophomores enrolled in an introductory psychology course. They were obligated to participate in the study as a course requirement. All subjects were first tested in a group session and then individually. In the group session, a modified form of the Remote Associates Test (RAT) was administered (Mednick, 1962). This test is said to yield a measure of creative thinking and seems manifestly to demand a certain flexibility of associative processes. In the context of the present study, it seemed a suitable measure of conceptualization and could serve to evaluate intellectual functioning of the kind which is hypothetically vulnerable to ineffective anal defenses. Its specially adapted form consisted of one part of 10, and a second part of 20 items. The first 10 required neutral solutions; the remaining 20 items were mixed, 10 requiring neutral and 10, anal solutions, appearing in random order. Both sets of neutral items were matched for difficulty. The instruction to the test was as follows: "In this test you will be presented with three words and asked to give the word *related* to *all three*." For example, "sweet" is the solution word for COOKIES, SIXTEEN, and HEART. No anal illustration item was, however, presented. A sample anal item is SHOOT, HORSE, HOUSE —SHIT (SHOOT THE SHIT, HORSESHIT, SHITHOUSE).

Fifteen minutes were allowed for the first 10 items and 30 for the other 20. The test was scored for the number of items solved correctly.

The second task administered in the group session was a specially prepared questionnaire which contained 18 items tapping a variety of anal anxieties, 16 items tapping anal character traits, and 9 items tapping nonanal anxieties. Twenty-five neutral items were added to these. Of the anal character traits and anxiety items, about half dealt directly with matters of the toilet or feelings about dirt and cleanliness, and half with more derived traits and preoccupations commonly associated with anal character structure, according to the clinical literature (e.g., Reich, 1949). Examples of anal anxiety items are: "Handling wet and slimy things upsets me," and "I get upset when I realize how much time I waste that should be used for study." Examples of anal character traits are: "I bathe (or shower) frequently," and "I am a punctual person." An example of a nonanal anxiety is: "I fear that I am mentally ill." The subjects endorsed each item on an 11-point scale from "very characteristic of me" to "not characteristic of me." In this way, every subject earned scores for anal anxiety, anal character, and nonanal anxiety. Thus, in addition to the measure of adaptive strength drawn from the action test, the questionnaire provided the second common criterion of defense effectiveness, freedom from anxiety.[2]

---

[2] Further factor-analytic information concerning this questionnaire may be obtained from Schlesinger (1963).

The individual session was begun with a Dot Estimation task. This test was administered to evaluate the subject's decisiveness in making simple judgments. Fifteen cards, 8 × 11½ inches in size, with random arrays of dots, from 19 to 267 in number, were exposed to the subject for 2 seconds at the rate of one card every 5 seconds. During the 3 seconds intervening between exposures he was to write down his "estimate of the number of exposed dots." In a second trial, the subject was to repeat the procedure with the difference that he himself would determine when the next card was to be exposed. He was informed that in this trial his "score would be equally weighted for accuracy and speed of estimation."

Subsequent to this task, the action test assessing the effectiveness of defenses against anal stimulation was administered. The subject's task required that he immerse his arm, up to the elbow, into a bucket filled with water at room temperature which was hidden from his view by a curtain. Three irregularly shaped flat bits of aluminum, larger than those commonly used in jigsaw puzzles, had been dropped by the experimenter to the bottom of the bucket. A card with the outlines of six such pieces was displayed in full view of the subject. The instruction was to explore the contours of the three pieces, one after another, and to decide as quickly as possible to which of the displayed outlines it corresponded. The subject was required to leave his arm immersed until he had reached a decision. Then he was to remove the piece and place it on the table beside the bucket so that the experimenter might record the results. Four trials of three judgments each were administered using water as a medium. This was followed by four trials in a mixture of used crankcase oil and two pounds of flour. This dirty and odorous medium was assumed to arouse anxieties and other responses similar to those commonly attached to feces. While there were marked individual differences in response to this task—some subjects had had experience working in gas stations—it was quite clear from most subjects' behavior and verbal report that they found it decidedly unpleasant. Different aluminum figures were used in the water and oil trials. The number of correct judgments and the time required for each trial were recorded. To keep up their interest, subjects were informed of the number of correct judgments made in each trial.

RESULTS

Each subject's performance on the action test was expressed in quotients of time-required-per-correct-matching in water and in oil. A subject whose efficiency in oil was less than it had been in water was said to be defending relatively ineffectively against the presumed anal stimulation of the oil medium. However, since the entire sample showed an average increase in efficiency from water to oil, the subjects were divided at the median of efficiency shift scores (the difference between water and oil efficiency). Those below the median are referred to as good defenders (GD), implying that they either had no noteworthy conflict about anal stimulation or were dealing effectively with whatever conflict they had. Those above the median are referred to as poor defenders (PD).

The prediction that ineffective defenses against anal stimulation would be associated with indecisiveness was borne out by a sign-test evaluating the estimation latencies for Part II of the Dot task. On 11 out of the 15 stimulus cards the PDs' latencies were longer than the GDs' ($p = .006$). There are no consistent differences in accuracy of dot estimation between GD and PD subjects.

In order to test the prediction that ineffective defenses against anal stimulation would be associated with an intellectual disturbance, the relationship between action test and RAT performance was examined. The product-moment correlation between number of anal items solved and efficiency shift scores on the action test was $r = -.29$ ($p = .05$) indicating that GDs tend to solve more such items. But the correlation of action test with neutral RAT solutions was $r = -.01$. That anal and neutral items tap fairly distinct psychological functions is further suggested in that the number of anal RAT solutions was only slightly correlated with the number of neutral ones ($r = .13$).

It was a matter of interest to compare performance on the newly developed action test with questionnaire ratings, a more conventional instrument for the assessment of character traits and anxieties. Whether a measure of performance which did not aim directly at measuring the conscious experience of anxiety yielded the same results as a measure based on the subject's report of such manifest anxiety, would also be of theoretical interest. Accordingly, two groups of 12 subjects each were randomly chosen from the extremes of the anxiety questionnaire. The range of anxiety scores for the high-anxious group was from 14 to 7 and that for the low-anxious group from 0 to 3 with a possible maximum score of 18. These two groups performed differently on the action test, high anal-anxious scorers showing deterioration in the oil medium, and low scorers showing improvement ($t = 2.06$, $p < .05$). This indicates that subjects who describe themselves as *anxious* about anal concerns, show relative impairment of performance when the water medium is replaced by the oil medium. Neither anal *character* nor nonanal anxiety showed such an association with the action test. For each of these latter comparisons, $r = .02$. The differential association of anal anxiety and character traits with the performance criterion is quite in keeping with theoretically grounded expectations and will be discussed below. The questionnaire measures of anal anxiety and anal character were, however, significantly correlated with each other ($r = .32$, $p < .05$).

Would the anal anxiety questionnaire have served as well as the action test in confirming the predictions about indecisiveness and intellectual disturbance? To deal with this question, parallel computations were made. The measures of anal anxiety drawn from the questionnaire were positively correlated with the Dot latencies. On 13 of the 15 stimulus cards, the high-anxious subjects had longer average latencies than the low-anxious ones ($p = .002$). However, it was not found that they solved more anal RAT items, as was the case in comparing GDs and PDs. The action measure, therefore, yielded a correlation that would not have been obtained from the questionnaire.

DISCUSSION

Performance of the action test, as a sample of the subject's adaptive potential in response to anal stimuli, was expected to hinge on the effectiveness of his defense, not on the relative importance of anal eroticism in his personality make-up. This expectancy is borne out in that stable character traits commonly thought to be derived from anal interests were not correlated with efficiency shift scores while self-reported *anxiety* about anal impulses did predict these scores. Both anxiety and inadequate adaptation are generally deemed to be consequences of ineffective defenses. Since anal character traits are, however, also known to occur in more or less stable people, free of anxiety, one would expect a far from perfect correlation between anal character and anal anxiety in the population at large. An underlying assumption, as yet to be specifically tested, is that while anal anxiety ought to be found only when evidence of anal character traits is also present, it is quite common to find the latter without the former. The statistical correlation should therefore be moderate, as it was in the present sample ($r = .32$). This permits the interpretation that there may have been well-functioning anal personalities among the GDs. From a theoretical standpoint, the distinction between having no anal conflicts and having anal conflicts, but resolving them successfully, is dubious. In principle, it is impossible to avoid dealing with erotic and aggressive impulses in oneself. The above distinction, therefore, refers mainly to the degree in which the successful resolution is evident to an observer or so deeply repressed or fully integrated as to escape notice.

Although the reported findings seem to bear out clinical impressions of the obsessive-compulsive personality constellation, especially of the linkage between an involvement with dirt and a characterological emphasis on intellectualization and action inhibition, the findings can be explained in at least two ways. The first explanation is based on the presumed causal sequence of insufficient or undependable defensive efforts failing to suppress the mounting anxiety which is triggered by unacceptable anal impulses. This leads to functional disruption. A more detailed account of this is given elsewhere (Rosenwald, 1961).

A second explanation is possible. The PDs' relative impairment in the oil medium may not have resulted from anxiety, but from a playful lingering over a substance which, in the case of these subjects, was enjoyable precisely because it lends itself to implicit anal experiences or fantasies. In other words, a kind of self-indulgence, rather than anxiety, may have retarded their performance. This would be consistent with the clinical understanding of anal, or compulsive, personalities. Even severely neurotic individuals who observe stringent taboos against anal temptations frequently betray a fascination with dirt and destructiveness. However, in the present context, playfulness or fascination constituted an adaptive weakness because all subjects knew that they were being timed. The instructions to the action test as well as to the previously administered Dot task had set a competitive pace for the entire session. Furthermore, that the action test was positively related to the anxiety questionnaire lends additional credence to the broad maladaptation hypothesis, whether the inefficiency in the oil medium is interpreted as resulting from functional disruption, playful lingering, or a combination of both these factors.

The RAT findings also suggest that sensitization factors, that is, factors of fantasy or other symbolic activity, contributed to the action test performance. The negative correlation between anal RAT solutions and efficiency shift scores may be explained by reference to the PDs' sensitivities to anal stimuli: Just as they were captivated and/or hindered by the anal appeal of the oil medium, so they were also alert to the opportunities for a symbolic anal response to the relevant RAT items. In other words, neither sensitization nor disruption through anxiety can be ruled out from the available data. The concatenation of erotic and anxiety factors is a frequently observed clinical phenomenon and can be conceived as an outcome of ineffective defenses.

Several ambiguities arise in considering each of the employed measures separately. For instance, the drop in efficiency from the water to the oil medium may reflect a generalized discomfort reaction to unfamiliar stimuli or an aspect of tactile insensitivity rather than a functional disruption caused by attitudes toward dirt. It is also possible that Dot Estimation scores are related to psychophysical acuity. It may, therefore, be hypothesized that subjects whose sensory receptivity for tactile cues is easily disturbed also show a judgmental distractibility in response to visual cues. That is to say, subjects who were disturbed in one situation were also disturbed in another. Similarly, it is possible that the correlation between anal anxiety scores on the questionnaire and efficiency shift scores is not mediated by particular impulse defense patterns. All one need assume is that a subject will respond accurately when asked whether he likes or dislikes a mess. Those who say they dislike it, prove to be consistent in that they perform poorly under messy conditions. Finally, one can dispense with the defense theory in explaining the PDs' advantage on anal RAT items. One can say that those subjects who had the greatest tolerance for the dirty medium, that is, the least squeamish subjects, also have fewer inhibitions against somewhat improper solution words.

However, when considered simultaneously, these various reinterpretations are not so persuasive. Taken together, they would imply that the action test is at once a measure of the subject's distractibility under the stress of unfamiliar conditions, of the subject's squeamishness about impropriety, and of the subject's positive liking for messiness. Although one cannot rule out that sensory acuity (and tactile sensitivity specifically) affects performance on the action test, this alone could not account for the correlation of the action test with the RAT, with the questionnaire, or perhaps even with the Dot task. Furthermore, while the action test and the Dot task have a sensory dimension and a concentration factor in common, this is not the case for the anal anxiety scores on the questionnaire, which were related both to the action test and to the Dot Estimation scores. All in all, it seems more parsimonious to abide with the theory and with the clinical observations which prompted this experiment. But further research linking the action test and questionnaire with factors like stubbornness, emotional constriction, and stinginess, will perhaps provide a more cogent demonstration of the cluster of anal traits.

Although the RAT is supposedly a measure of creative thought, the superior performance of the PDs may be viewed apart from considerations of creativity because their advantage was confined to anal items. The lack of correlation between these and the neutral RAT items suggests that solution of anal items is not so much a sign of creativity, as of a readiness for the expression of anal symbols and themes. On the other hand, the PDs' intellectual disturbance did not cause them any disadvantage in comparison with GDs' RAT performance.

In the past, researchers have often assumed that an emotional conflict and the resulting defensive efforts necessarily entail the eclipse of the relevant symbols and ideas from the subject's mind. However, especially in the case of anal characters, clinicians often remark on a rather eager, though incomplete, expression of libidinally or aggressively charged ideas. Defenses such as projection, reaction formation, and isolation and a host of so-called counterdefenses avert unconscious psychic dangers by modifying the content of forbidden wishes, rather than by eradicating them altogether. This phenomenon is reminiscent of a recent study in which lowered perceptual thresholds were observed in subjects with a preference for "externalizing" defenses (Lewit, Brayer, & Leiman, 1962).

An extreme form of intellectual disturbance found in compulsion neurosis is the emergence into awareness of unacceptable dirty or destructive ideas. Further research may show whether the present RAT finding is a mild form of such a disturbance. Because particular defenses are associated with particular contents, great care must be taken in generalizing the present RAT results. For instance, ineffective resolution of conflicts about dependency might be associated with predominantly avoidant defenses rather than counterdefenses. The result might be a suppression of, rather than a sensitization to, pertinent symbols and ideas. In brief, the hypothesis concerning intellectual disturbance in PDs is supported inasmuch as they show a bias which sensitizes them to drive-related, but not to neutral, associative contents.

REFERENCES

BARNES, C. A statistical study of the Freudian theory of levels of psychosexual development. *Genetic Psychology Monographs,* 1952, **45**, 115–175.

FREUD, A. *The ego and the mechanisms of defense.* New York: International Universities Press, 1946 (orig. pub. 1936).

LEWIT, D. W., BRAYER, A. R., & LEIMAN, A. H. Externalization in perceptual defense. *Journal of Abnormal and Social Psychology,* 1962, **65**, 6–13.

MEDNICK, S. A. The associative basis of the creative process. *Psychological Review,* 1962, **69**, 220–232.

REICH, W. *Character-analysis.* New York: Orgone Institute Press, 1949.

ROSENWALD, G. C. The assessment of anxiety in psychological experimentation: A theoretical reformulation and test. *Journal of Abnormal and Social Psychology,* 1961, **62**, 666–673.

SCHLESINGER, V. J. Anal personality traits and occupational choice: A study of accountants, chemical engineers, and educational psychologists. Unpublished doctoral dissertation, University of Michigan, 1963.

SEARS, R. Studies of projection: I. Attribution of traits. *Journal of Social Psychology,* 1936, **7**, 151–163.

(Received July 22, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 4, 310–314

CORTICAL INHIBITION, DRIVE LEVEL, AND CONDITIONING[1]

P. O. DAVIDSON
University of Calgary
R. W. PAYNE AND R. B. SLOANE
Temple University Medical School

The performance of 40 neurotic female patients on finger withdrawal and GSR conditioning to shock was correlated with several personality and objective measures of drive level and cortical inhibition. All correlations between conditioning indices for the 2 responses were positive and significant ($p < .05$) for this group. The correlations of the drive measures with both conditioned responses were all consistent with Spence's (1958) drive theory although only 3 of these correlations were significant. There was no consistent relationship demonstrated between measures of cortical inhibition and conditioning. The implications of these and other results for the hypothesis of a general factor of conditionability were discussed.

In a previous study using normal university students (Davidson, Payne, & Sloane, 1964), conditioning indices for finger withdrawal and GSR were nonsignificantly correlated with measures of introversion, neuroticism, and manifest anxiety. Furthermore, there were no significant correlations between the two conditioning procedures for this group.

While these results may not refute Eysenck's (1957) theory of a relationship between cortical inhibition and conditioning, because of limiting conditions he has recently specified (Eysenck, 1965), they did fail to confirm the notion implicit to his theory of a general factor of conditionability (Eysenck, 1957, 1960b). These results also failed to support Spence's (1958) theory of the relationship of drive (manifest anxiety) and conditioning.

Failure to find significant results in studies relating personality variables to conditioning (cf. Willett, 1960) may sometimes be due to the use of very homogeneous samples (Franks, 1961). Since the previous study had used a large sample of student volunteers there could have been some restriction in the ranges of their test scores. For the present study, a sample was chosen from a neurotic, rather than a normal, population in order to obtain subjects with extreme scores on tests of anxiety, extraversion, and neuroticism. It could be the case, as suggested in the previous study, that the relationship between these variables and conditioning is not linear and homoscedastic. If this were so then there might be a correlation between conditionability and personality measures when—with respect to these measures—an extreme group is tested instead of an unselected group.

Spence and his colleagues have relied heavily on a personality scale of manifest anxiety (Taylor, 1953) as the measure of drive level in their studies of performance in the eyelid conditioning situation (cf. Spence & Taylor, 1951; Taylor, 1951). However, Spence (1958) has noted that other measures of drive would be equally suitable. Specifically, the Neuroticism scale (N) of the Maudsley Personality Inventory (MPI; Eysenck, 1959) has been suggested as a good measure of drive by Eysenck (1960a) and adopted by Spence and Spence (1964). Malmo (1957) has suggested that autonomic measures, such as baseline GSR (B.L. GSR), can also serve as useful indices of drive level.

Eysenck and his colleagues have preferred to use the Extraversion scale (E) of the MPI as a measure of cortical inhibition when testing predictions from his theory. Eysenck (1957) suggested, as an alternate measure of cortical inhibition, the sedation or sleep threshold developed by Shagass (1954, 1956). This is a measure, in milligram per

---

[1] This research was carried out at Queen's University, Kingston, Ontario, with the assistance of Canadian Dominion Provincial Mental Health Grant No. 605-5-318.

TABLE 1

CORRELATIONS BETWEEN PERSONALITY AND CONDITIONING

Variable	Personality measures						Conditioning indices			
	Drive level			Cortical inhibition			Finger withdrawal		GSR	
	MA	N	B.L.GSR[a]	E	Sed.T.	S.F.GSR	T.T.	E.T.	T.T.	E.T.
MA		.76**	.14	−.29	.00	−.26	.30*	.29	.23	.23
N			.17	−.29	.04	−.20	.13	.14	.21	.27
B.L.GSR[a]				−.16	−.13	−.05	.15	.06	.43**	.33*
E					−.57**	.12	.12	.06	−.17	−.22
Sed.T.						.35*	−.29	−.32*	.16	.09
S.F.GSR							.05	−.03	−.07	−.02

Note.—$N = 40$.

[a] B.L.GSR is reflected since a low baseline is supposed to relate to high drive.

* $p < .05$.

** $p < .01$.

kilogram body weight, of the amount of sodium amytal needed to produce sleep. While the neurophysiological and biochemical mechanisms underlying this test may not be fully understood, it has been suggested as a useful tool in differentiating introverts and extroverts (Eysenck, 1961; Shagass & Naiman, 1956). A further measure of cortical inhibition has been proposed by Greiner and Burch (1955) who found that the number of spontaneous fluctuations in GSR (S.F. GSR) showed a roughly linear relationship to the degree of central inhibition or arousal that could be obtained using depressant and stimulant drugs. Silverman, Cohen, and Shmavonian (1959) have also indicated S.F. GSR as a good index of central inhibition.

The above measures were incorporated into the present study.

METHOD

Subjects

The subjects were 40 female neurotic patients admitted to the psychiatric ward of the Kingston General Hospital, Ontario. The subjects were accepted in their order of admission provided they were between the ages of 16 and 65 years. Attempts were made to exclude subjects who had been on a high dosage of drugs and all subjects were routinely kept off any medication except night sedation for 3 days prior to testing. The mean age of this group was 35 ± 13 years and their mean IQ was 93 ± 12 as assessed by the Mill Hill Vocabulary Test (Raven, 1958).

Apparatus and Method of Recording

The apparatus and procedure were the same as those described in the previous investigation (Davidson et al., 1964). The conditioned stimulus (CS) was a 750 cps tone of 200 milliseconds duration followed 400 milliseconds later by a UCS of 200 milliseconds duration. The UCS, an unavoidable dc electric shock to the right-middle finger, was maintained at the subjects' threshold of reaction to pain. The mean UCS strength for this group was 2.2 ± 1.2 milliamperes. The responses to be conditioned—finger withdrawal and GSR to shock—were recorded on a Grass Model 5 polygraph simultaneously with measures of heart rate, blood pressure, respiration, and finger plethysmography.

Procedure

The subject lay on a bed in a sound-deadened electrostatically shielded and air-conditioned room. After the recording electrodes had been attached the subject was instructed to lie quietly, as relaxed as possible, for 15 or 20 minutes. The measure of B.L. GSR was obtained after at least 10 minutes rest and before any testing. S.F. GSR was scored as the number of nonspecific fluctuations from baseline which exceeded 1,000 ohms amplitude during a 5-minute resting period.

After adjusting the UCS strength the subjects were instructed regarding the nature of the conditioning procedure and cautioned not to respond voluntarily. Each subject was then given 50 acquisition trials with one test trial of tone alone randomly interspersed among every five acquisition trials. The conditioning procedure was followed immediately by 20 extinction trials using tone alone.

GSR and finger withdrawal were conditioned simultaneously. Two indices of finger withdrawal conditioning were taken; the number of responses to the 10 test trials in the acquisition series, and the number of responses during the 20 extinction trials. The same indices were used for the GSR conditioning. The scoring criteria suggested by Becker (1960) were used in counting GSRs to the CS as conditioned responses.

Prior to all testing the subjects had been given the Manifest Anxiety (MA) scale and the MPI.

Subsequent to the conditioning procedures, sedation thresholds (Sed. T.) were determined by a method described in detail by Claridge and Herrington (1960).

RESULTS

Correlations among the Personality Measures

The correlation matrix for the personality measures is shown in Table 1.

As was expected, a statistically significant relationship was found between two of the measures of drive level (MA and N). The correlations of the third measure of drive (B.L. GSR) with the other two, while positive, were not significantly different from zero. The intercorrelations of the three measures of cortical inhibition were in somewhat closer agreement—with two out of the three showing a statistically significant relationship. Sed. T. correlated in the direction predicted since high Sed. T. should relate to low E scores (Shagass & Naiman, 1956).

None of the drive level measurements correlated significantly with any of the measurements of cortical inhibition, although two of the three correlations with the E scale were just short of significance ($p < .07$). A negative relationship of this order between extroversion and neuroticism (drive level) is often obtained when an extreme group such as the present one is tested (cf. Eysenck, 1959).

Correlations with Conditioning

The conditioning indices were the number of responses given during the test trials (T.T.) and extinction trials (E.T.). The correlations between these indices for the two conditioned responses and each of the personality measures are shown in Table 1. While the results shown in this part of the table were not conclusive, some relationships did emerge. All the correlations of the drive measures with the conditioning indices were positive; three were statistically significant and five more were close to significance. The most consistent relationships seemed to be with the MA scale.

The high correlations for B.L. GSR with GSR conditioning may be unique to that response. The possibility of these correlations being spurious due to the known relationship between baseline level and response amplitude of the GSR (Martin, 1960) is not very plausible. This spurious correlation may result because GSRs must show a certain minimum amplitude in the allowable latency period in order to be counted as conditioned responses (Becker, 1960). The amplitude of a GSR relates to the baseline level according to the Law of Initial Value (Wilder, 1957). The GSR—a drop in resistance—can show a larger amplitude when there is a high baseline from which to drop. The Law of Initial Value would therefore suggest a spurious correlation (due to false positive *CR*s) between *high* baseline and GSR conditioning. The results given in Table 1 actually showed a significant correlation between *low* baseline and GSR conditioning (B.L. GSR correlations were reflected in the table). It would seem far more likely that false positive *CR*s due to a high rate of spontaneous fluctuations in GSR would produce spurious correlations with GSR conditioning but inspection of Table 1 showed that this was not the case.

TABLE 2

INTERCORRELATIONS OF CONDITIONING INDICES

Variable	Conditioning indices			
	Finger withdrawal		GSR	
	T.T.	E.T.	T.T.	E.T.
Finger withdrawal				
T.T.		.71**	.37*	.30*
E.T.			.42**	.47**
GSR				
T.T.				.73**
E.T.				

Note.—N = 40.
* $p < .05$.
** $p < .01$.

The relationships between measures of cortical inhibition and conditioning did not show the same consistency of results that were shown by the drive measures. Seven of these correlations were negative and five were positive. The two highest correlations (of sedation threshold with conditioning) were in the opposite direction to that predicted by Eysenck's theory.

Intercorrelations of Conditioning Indices

The correlations between acquisition and extinction of the conditioned finger withdrawal response and the conditioned GSR are shown in Table 2.

The correlation between the number of *CR*s during the test trials and extinction trials was highly significant for both finger withdrawal and GSR. The previous study by Davidson et al. (1964) had also shown this result but in that study there were no significant correlations between the two responses. In the present study, as shown in Table 2, all four of these correlations were positive and significant.

DISCUSSION

In comparing this study with the previous one several interesting findings emerged. Using normal subjects conditioned with an avoidance procedure no generality of conditionability was evidenced. A small sample of normal subjects conditioned with a nonavoidance procedure also showed no significant correlations between the two conditioned responses. However, a group of neurotics conditioned with a nonavoidance procedure did show evidence of a generality of conditioning. These results—to the extent that they are replicable using other responses—suggest that demonstrating a generality of conditioning may depend upon the kind of sample studied. If neurotics show a generalization of conditionability while normals do not, this might explain why neurotic individuals show (according to Mowrer, 1950; and Dollard & Miller, 1950) such a profusion of nonadaptive conditioned responses. Extrapolating from our results it would seem that normal persons experiencing an unavoidable and unpleasant stimulus situation may develop *either* a conditioned withdrawal response or a conditioned autonomic response to that situation, whereas neurotics develop *both* a conditioned withdrawal response and a conditioned autonomic response.

The results from this study support Spence's theory of the relationship of drive level to conditioning but, considering this study in conjunction with the previous one, it may not be the simple linear relationship that Spence assumes. The results from this study are not compatible with Eysenck's theory relating cortical inhibition to conditioning because the only significant correlations in this respect were opposite in direction to his prediction. They do give partial support to Eysenck's hypothesis of a generality of conditionability but with the limitation that this finding may only appear in neurotic subjects.

REFERENCES

BECKER, W. C. Cortical inhibition and extraversion-introversion. *Journal of Abnormal and Social Psychology,* 1960, **61**, 52–66.

CLARIDGE, G. S., & HERRINGTON, R. N. Sedation threshold, personality, and the theory of neurosis. *Journal of Mental Science,* 1960, **106**, 1568–1583.

DAVIDSON, P. O., PAYNE, R. W., & SLOANE, R. B. Introversion, neuroticism, and conditioning. *Journal of Abnormal and Social Psychology,* 1964, **68**, 136–143.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

EYSENCK, H. J. *The dynamics of anxiety and hysteria.* London: Routledge & Keegan Paul, 1957.

EYSENCK, H. J. *The Maudsley Personality Inventory.* London: University of London Press, 1959.

EYSENCK, H. J. A factor analysis of selected tests. In H. J. Eysenck (Ed.), *Experiments in personality.* Vol. 2. New York: Macmillan, 1960. Pp. 234–244. (a)

EYSENCK, H. J. The place of theory in psychology. In H. J. Eysenck (Ed.), *Experiments in personality.* Vol. 2. New York: Macmillan, 1960. Pp. 303–315. (b)

EYSENCK, H. J. The effects of drugs on behavior. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961. Pp. 634–696.

EYSENCK, H. J. Extraversion and the acquisition of eyeblink and GSR conditioned responses. *Psychological Bulletin,* 1965, **63**, 258–270.

FRANCKS, C. M. Conditioning and abnormal behavior. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961. Pp. 457–487.

GREINER, T. H., & BURCH, N. R. Response of human GSR to drugs that influence the reticular formation of brain stem. *Federation Proceedings,* 1955, **14**, 346.

MALMO, R. B. Anxiety and behavioral arousal. *Psychological Review,* 1957, **64**, 276–289.

MARTIN, I. Skin resistance and relationship to GSR conditioning. *Journal of Mental Science,* 1960, **106**, 281–287.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.

RAVEN, J. C. *Guide to using the Mill Hill Vocabulary Scale with the Progressive Matrices Scale.* London: Lewis, 1958.

SHAGASS, C. The sedation threshold. A method for estimating tension in psychiatric patients. *E.E.G. and Clinical Neurophysiology,* 1954, **6**, 221–223.

SHAGASS, C. Sedation threshold. A neurophysiological tool for psychosomatic research. *Psychosomatic Medicine,* 1956, **18**, 410–419.

SHAGASS, C., & NAIMAN, J. The sedation threshold as an objective index of manifest anxiety in psychoneurosis. *Journal of Psychosomatic Research,* 1956, **1**, 49–57.

SILVERMAN, A. J., COHEN, S. I., & SHMAVONIAN, B. M. Investigation of physiological relationships with skin resistance measures. *Journal of Psychosomatic Research,* 1959, **14**, 65–87.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *American Psychologist,* 1958, **13**, 131–141.

SPENCE, K. W., & SPENCE, J. T. Relation of eyelid conditioning to manifest anxiety, extraversion and rigidity. *Journal of Abnormal and Social Psychology,* 1964, **68**, 144–149.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *Journal of Experimental Psychology,* 1951, **42**, 183–188.

TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *Journal of Experimental Psychology,* 1951, **41**, 81–92.

TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology,* 1953, **48**, 285–290.

WILDER, J. The law of initial value in neurology and psychiatry. *Journal of Nervous and Mental Diseases,* 1957, **125**, 73–86.

WILLETT, R. A. Measures of learning and conditioning. In H. J. Eysenck (Ed.), *Experiments in personality.* Vol. 2. New York: Macmillan, 1960. Pp. 157–192.

(Received September 10, 1965)

ERRATUM

In the article, "Stress Situations, Birth Order, and Psychological Symptoms," by Barbara Snell Dohrenwend and Bruce P. Dohrenwend in the June 1966 issue, the first sentence of the first paragraph on page 222 should read: "Turning again to the relatively high rate of symptoms among later-born subjects of disadvantaged ethnic status, and following the lead suggested above that these symptoms are reactions to contemporary events, the investigators attempted to find conditions in the subjects' current situations which would explain their symptoms in terms of the adaptive mechanisms described by Schachter (1959) and by Wrightsman (1960)."

Volume 71 October 1966 Number 5

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute and Columbia University

KURT SALZINGER
New York State Psychiatric Institute and Brooklyn Polytechnic Institute

JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR — Managing Editor
MARY C. MCGARRAGHY — Editorial Assistant
HENRY S. SALMON — Advertising and Business Manager
ROBERT D. GORMAN — Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1966 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 315–327

ASSOCIATION, ABSTRACTION, AND THE CONCEPTUAL ORGANIZATION OF RECALL:

IMPLICATIONS FOR CLINICAL TESTS[1]

A. E. WILLNER AND W. E. REITZ[2]
Menninger Foundation

Recent discoveries about associative meaning were hypothesized to have important implications for the problem of assessment of abstraction by clinical tests. The major hypothesis was that association functions in abstraction tests by eliciting both (a) the concept required for the abstraction item, and (b) the associative neighborhood containing the required concept. A series of 4 experiments investigated this hypothesis for both the WAIS Similarities and BRL Object Sorting Tests. In a 5th experiment, an abstraction test (the BRL) was administered as a memory task in order to investigate the role of association in the conceptual organization of recall. All experiments provided strong support for the major hypothesis and have implications for construction of abstraction tests of enhanced clinical sensitivity.

The recent rapid increase in knowledge about associative meaning has produced valuable returns in the study of verbal recall, including more accurate prediction of the amount and organization of recall (cf. Cofer, 1964; Underwood & Schulz, 1960). The present authors contend that this development has important implications for clinical psychology, since an increased understanding of the role of associative processes in the process of abstraction could lead to construction of more effective clinical tests of abstraction. The experiments reported here aim at an improved understanding of the process of abstraction as assessed in clinical tests. The general hypothesis of this research is that association is an essential process in the total sequence of conceptual thought. Association is conceived of as a logical, although not necessarily chronological, first step in the process of abstraction. It is hypothesized to function in abstraction both by directly eliciting the required concept and also by eliciting the "associative neighborhood" containing the required concept. The term *associative neighborhood* refers to a cluster of words and ideas organized by similarity in associative meaning. Similar concepts have recently been formulated in psycholinguistics and psychology.

Whorf (1956) theorized that the basic human concepts underlying all languages were organized into "neighborhoods" by a special type of association which he called "connections." Scheerer (1959) used the term "spheres of meaning" to explain errors, made by schizophrenic and aphasic patients, which were associatively close to the correct answer. Deese (1962), in discussing the role of associative meaning in concepts, said that the closer the relation between two words, the more likely they were to appear in the same "verbal environment." Pollio (1964) used the concept of "associative clusters . . . which evoke essentially similar connotative meaning responses [p. 200]" to explain alterations in the rate of association in a continuous association task. The concept, used by the above authors, of a cluster of words or ideas with strong associative interconnections will here be called an "associative neighborhood."

The major hypothesis investigated in this research is that association functions in abstraction by eliciting both (*a*) the required concept directly, and (*b*) the associative neighborhood containing the required concept. Several recent studies (Coleman, 1964; Freedman & Mednick, 1958; Mednick & Halpern, 1962; Underwood & Richardson, 1956) have demonstrated that the difficulty of learning a new concept is a function of the strength of the associative link between the concept and the stimulus words. Thus, the concept is more easily learned if it is a dominant rather than a nondominant associate of the stimulus words. The concept-formation tests used in these studies, however, were all especially constructed to test the hypothesis, and the present authors could find no study applying these methods to familiar, standard tests of abstraction. These studies also did not investigate the more general case of two stimulus words eliciting the same association (regardless of whether it was the required concept), but were restricted instead to instances where the required concept was an associate of the two stimulus words. Investigation of the more general case seems quite important because it may indicate that the required associative neighborhood was elicited, although the required concept was not produced.

---

[1] This research was supported by a grant (MH 11316 01) from the National Institute of Mental Health, United States Public Health Service. The research was also supported in part by a Postdoctoral Clinical Research Fellowship to the second author from the National Institute of Mental Health.

[2] Willard Reitz is now at the University of Western Ontario. The authors are indebted to Bay Chotlos, John Lebbert, and Dixie Donnelly for their help in collecting and tabulating the data for this experiment.

Association Variables

The effect of association upon difficulty of an abstraction task, for example, a similarities test item, may be examined in several ways. Assuming that each of the two stimuli elicits a network of associations, one may investigate the degree to which the two networks overlap. The degree of associative overlap (which will be called *mutual association*) is assumed to indicate how closely the two stimuli are located in an associative neighborhood. One may also, as in the four studies just cited, investigate whether the associations elicited by the stimuli include the "correct concept" required for the test item (this will be called *stimulus-concept* (S-C) *association*).

The research reported here investigates the effect of both S-C and mutual association upon the difficulty level of two types of standard abstraction tests: The WAIS Similarities Test and an object sorting test, the BRL (Bolles, Rosen, & Landis, 1938). The BRL was selected because it has a quantitative scoring system (Rapaport, 1945). This research consists of a series of five experiments. Similarities Experiments I and II investigate the effect of association upon abstraction in the WAIS Similarities Test. Object Sorting Experiments I and II investigate the effect of association upon abstraction in the BRL Object Sorting Test. Object Sorting Experiment III investigates the effect of association upon the conceptual organization of recall in the BRL.

The data for these experiments were gathered in the following order: (*a*) Similarities Experiment I and Object Sorting Experiment I; (*b*) Similarities Experiment II and Object Sorting Experiment II; (*c*) Object Sorting Experiment III.

Similarities Experiment I

Method

The stimulus materials were two sets of 13 words, the first and last words, respectively, of the 13 WAIS Similarities Test items. Data for this experiment, and for Object Sorting Experiment I, were gathered at the same time, from the same subjects (*Ss*). The *Ss*, 46 introductory psychology students from Washburn University, were group tested in their classrooms, in two groups of approximately equal size. All *Ss* were asked to give three word associations to each word in the first set, and then three associations to each word in the second set. The *Ss* were asked to give three associations because (*a*) the first association given in an open-choice word-association task is often an opposite, which would not be helpful in a similarities task, and (*b*) three associations allow a more precise measure of the number of mutual associations than only one association.

Predictions

1. The required concept will more often be given as an S-C association for easy than for difficult Similarities Test items.
2. Mutual associations will more often be given for easy than for difficult Similarities items.

Results

The data used were the total number of "correct" scores for the three associations to each stimulus word. Correct scores were association responses which would be scored correct for the corresponding WAIS Similarities Test item. Since Similarities Test responses can be scored 2, 1, or 0, each correct concept was given the appropriate score according to the WAIS manual. The items were divided into three groups, including the four easiest, the five moderate, and the four most difficult items. Mean scores were used so that the unequal number of items in the three groups would not bias the results. The scores for both words of the item were pooled in the analysis. The data were treated by analysis of variance; Table 1 shows the results.

The variance component attributable to difficulty level of the items was highly significant ($p < .001$). The means were .80, .27, and .13, respectively, for easy, moderate, and difficult items. The Newman-Keuls test for a series of individual mean comparisons showed highly significant differences ($p < .001$) for easy versus moderate and difficult items, and ($p < .05$) for moderate versus difficult items. The correlation between correct scores for associations to Similarities items and the ranked order of difficulty of the items also provides a measure of this relationship. The obtained rank-difference correlation, $-.83$ ($p < .005$, one-tailed), lends further support to the hypothesis that S-C association has a significant effect upon difficulty level of Similarities items.

Prediction 2 was investigated by recording the number of mutual associations given to each of the 13 Similarities Test items. The items were then divided into three levels: easy, moderate, and difficult; and the mean number of mutual associations for items of the three levels of difficulty was analyzed by the Friedman two-way analysis of variance; the levels differed significantly ($p < .001$). The correlation between the number of mutual associations for each Similarities item and the difficulty level of each item provides another measure of the relationship. In this analysis, the mutual associations scored 2, 1, and 0 were treated separately. Rank-difference correlations were computed between the ranked order of similarities item difficulty and the number of associations scored 2, 1, and 0; these were, respectively, $-.79$ ($p < .01$), $-.61$ ($p < .05$), and $-.59$ ($p < .05$). Associations scored 2 or 1 are both mutual and S-C associations; those scored 0 are mutual but not S-C associations.

These results support the hypothesis that association functions in Similarities Test items by eliciting the required concept directly (S-C association) and also by eliciting the associative neighborhood containing the required concept (mutual association). The method used, however (presenting each word of the pair of words comprising the Similarities item separately to *S*s and asking for associations to each word), differed from the structure of the actual Similarities Test in which the words are presented in pairs.

TABLE 1

ANALYSIS OF VARIANCE FOR S-C ASSOCIATION "CORRECT" SCORES FOR WAIS SIMILARITIES TEST ITEMS FOR SINGLE WORD PRESENTATION

Source	*SS*	*df*	*F*
*S*s	7.39	45	1.64*
Difficulty	11.77	2	58.84**
Residual	9.00	90	
Total	28.16	137	

Newman-Keuls test

Items	$\bar{X}_1$	$\bar{X}_2$	*F*
Easy versus moderate	0.80	0.27	66.37**
Easy versus difficult	0.80	0.13	105.61**
Moderate versus difficult	0.27	0.13	4.54*

* $p < .05$.
** $p < .001$.

SIMILARITIES EXPERIMENT II

This experiment investigates the effect of presenting the words in the Similarities Test as pairs of words, rather than separately, as an association task. The hypothesis is that when the words are presented in pairs, each pair of words will tend to provide a mutual focus for associative responses, with a resulting increment in the frequency with which the required concept is given as an association to the stimulus words.

Method

The stimulus materials used were the 13 pairs of words in the WAIS Similarities Subtest. Data for this experiment and for Object Sorting Test Experiment II were gathered at the same time and from the same *S*s. The *S*s, 60 introductory psychology students from Washburn University, were tested in two groups of approximately equal size. All *S*s were asked to give three word associations to each pair of words of the Similarities Test.

Predictions

1. The required concept will more often be given as an association for easy than for difficult Similarities Test items.
2. The required concept will be given more often as an association when the two words comprising the Similarities Test item are presented *as a pair* than when each word is presented separately.

Results

The data used to test these predictions were the total number of "correct" scores for the three associations to each stimulus. Correct scores were association responses which would be scored correct for the corresponding WAIS Similarities Test items. Thus each of the three associates to a given word or pair of words was scored 2, 1, or 0 according to the WAIS manual. (The data for single word presentation were gathered in the preceding experiment.) Since the single word condition included twice as many stimuli as the pairs-of-words condition, mean scores were used so that the unequal number of associations would not bias the results.

Both predictions were then tested by a repeated measure analysis of variance, unweighted means solution. The unweighted mean solution was used because the number of *S*s in each condition was different, 46 and 60 *S*s, respectively. Table 2 summarizes this analysis. The variance component attributable to difficulty level is highly significant ($p < .001$), strongly supporting Prediction 1. The means were 1.37, .74, and .39, respectively, for the easy, moderate, and difficult items. Newman-Keuls tests for a series of individual mean comparisons showed highly significant differences ($p < .001$) for each of the three comparisons.

Prediction 2 was also strongly supported, as shown by the highly significant effect ($p < .001$) of the variance component at-

TABLE 2

ANALYSIS OF VARIANCE FOR S-C ASSOCIATION "CORRECT" SCORES FOR WAIS SIMILARITIES TEST ITEMS FOR TWO METHODS OF PRESENTATION

Source	*SS*	*df*	*F*
Method of presentation	14.67	1	33.14**
Between *S*s, same method	46.02	104	
Difficulty	37.39	2	106.66**
Method × Difficulty	1.31	2	3.73*
Pooled *S*s × Difficulty	36.46	208	0.18

Simple effects

Condition	*M*			*t* tests		
	Easy	Moderate	Difficult	Easy versus moderate	Easy versus difficult	Moderate versus difficult
Single words (a)	.80	.27	.13	6.08**	7.68**	1.61
Pairs of words (b)	1.37	.74	.39	8.27**	12.86**	4.59**
t test: a versus b	5.64**	4.65**	2.57*			

Newman-Keuls test

Items	$\bar{X}_1$	$\bar{X}_2$	*F*
Easy versus moderate	1.12	0.54	101.94**
Easy versus difficult	1.12	0.28	213.82**
Moderate versus difficult	0.54	0.28	20.49**

* $p < .02$.
** $p < .001$.

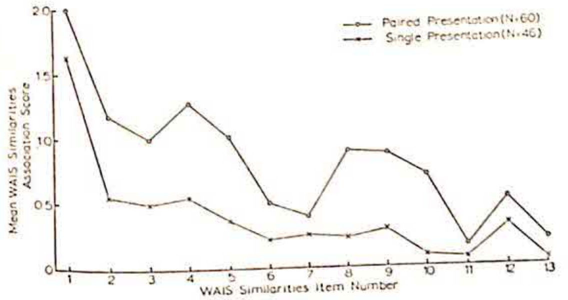


FIG. 1. Relationship between difficulty level and mean WAIS Similarities association score for single and paired presentations.

tributable to method of presentation (single word versus pairs). The means here were 5.11 and 10.72 for the single and paired presentations, respectively.

The significant interaction indicates that the effect of single versus paired presentation is not constant at all levels of difficulty. (Virtually all the simple effects comparisons are highly significant.) Plotting the mean WAIS Similarities Test score for each of the 13 items, and for single and paired presentation, revealed the effect to be relatively smaller at the extremes of difficulty and more pronounced at the intermediate level (Figure 1). A plausible explanation for this result is the following. At the easy level, the mean Similarities score was already so high as to leave little room for any increment, whereas at the difficult level the effect of association was so small that the presentation of the pair of words made little difference. At the intermediate level of difficulty, however, these "floor" and "ceiling" limiting effects were weakest, and presenting pairs of words could enhance the Similarities scores considerably. Finally, it should be pointed out that the paired condition resulted in higher Similarities mean scores for every one of the 13 items.

The correlation between correct scores for associations to Similarities items and ranked order of item difficulty provides another measure of this relationship. The obtained rank-difference correlation, $-.79$ ($p < .005$, one-tailed), suggests a high degree of relationship between the variables. This lends further credence to the hypothesis that association is significantly related to difficulty level of Similarities items.

OBJECT SORTING EXPERIMENT I

This experiment investigated whether the relation between association and abstraction demonstrated in the Similarities Test generalized to a less directly verbal test, the BRL Object Sorting Test. The BRL has two parts, an active and a passive sort. In the active sort, *S* groups each of 7 objects together with all the other (of 33 objects) that belong with it and explains why the objects belong together. In the passive sort, *S* is shown 12 groups of objects and asked, for each group, why the objects belong together. The 33 objects included in the test are the following: a large metal knife, a large metal fork and a large metal spoon, a small toy metal knife, a small toy metal fork and a small toy metal spoon, two sugar cubes, a red rubber sink stopper, a red cardboard circle, a white cardboard square, a white filing card, a pair of large metal pliers, a pair of small toy metal pliers, a large screwdriver, a small toy screwdriver, a toy hammer, a toy axe, two nails, a square block of wood with a nail in the center, a cigar, a rubber cigar, a cigarette, a candy cigarette, a red matchbook, a pipe, a red rubber ball, a lock, a bell, two corks, and a red rubber eraser. The active sort was investigated in this experiment.

Method

The stimulus materials consisted of six of the seven objects used in the active sort of the BRL (the first item was omitted because it is not standard). The *S*s, 46 introductory psychology students from Washburn University, were group tested in their classrooms. All *S*s were asked to give three word associations to each of the six objects.

Prediction

The required concept will more often be given as an S-C association for easy than for difficult items.

Results

Rapaport's (1945) division of the BRL items into three levels was not used as the criterion of difficulty level because it was based on the percentage of *S*s giving inadequate *sortings* (not *concepts*) for each item. The six items were instead divided, using Rapaport's raw data, into three levels (easy, moderate, and difficult items) on the basis of the concepts given by his *S*s. The *S*s' responses had previously been scored as "conceptual definitions," "functional definitions," "concrete," or "incorrect"; the corresponding responses were now assigned scores 3, 2, 1, and 0. The items were divided into the three levels of difficulty just mentioned on the basis of the mean score for each item.

The data used in this experiment were the total number of "correct" scores for the three associations to each stimulus object. Correct scores were association responses that would be scored correct for the corresponding BRL item. Mean scores were obtained for each group of items so that unequal numbers of items in each group would not bias the results.

Table 3 shows the results of the analysis of variance. The variance component attributable to difficulty level of the item was highly significant ($p < .001$), and the individual mean comparisons showed that the means for difficult versus easy and difficult versus moderate items were significantly different ($p < .001$), but no significant difference was found between means for easy versus moderate items.

Table 4 shows the percentage of correct responses to BRL active-sort items obtained from Rapaport's data and the corresponding percentage of correct association responses to the same stimulus-objects by *S*s in this experiment. In five of the six items there is no more than a 15% difference in level of accuracy between object sorting and association conditions.

TABLE 3

ANALYSIS OF VARIANCE FOR S-C ASSOCIATION "CORRECT" SCORES FOR BRL OBJECT SORTING TEST ITEMS

Source	*SS*	*df*	*F*
*S*s	103.43	45	1.54*
Difficulty	100.75	2	33.74**
Difficulty × *S*s	134.38	90	
Total	338.56	137	

Newman-Keuls test

Items	$\bar{X}_1$	$\bar{X}_2$	*F*
Easy versus moderate	1.93	2.29	1.95
Easy versus difficult	1.93	0.33	38.55**
Moderate versus difficult	2.29	0.33	57.86**

* $p < .05$.
** $p < .001$.

TABLE 4

RESPONSES TO STIMULUS OBJECTS FROM THE BRL ACTIVE SORT UNDER OBJECT SORTING AND ASSOCIATION CONDITIONS

Item	Object sorting norms percentage correct	Association data percentage correct
Fork	83	70
Pipe	100	85
Bell	11	15
Disc	76	63
Pliers	100	39
Ball	87	72

These data provide additional information about the effect of association upon concept formation, by showing similar results in quite different conditions: object sorting and object association. In the object sorting condition, Rapaport's *S*s selected a group of objects that "belong with" a given stimulus object and then explained the rationale for the grouping. In the association condition of this study *S*s did not know the task had any connection with object sorting and were asked only to give associations to the stimulus object. Despite the difference in conditions, similar percentages of "correct" responses were obtained when the rationales given by Rapaport's *S*s, and the associations of the present *S*s, were scored by Rapaport's criteria.

OBJECT SORTING TEST EXPERIMENT II

In the preceding experiment, it was not possible to assess the effect of mutual association upon the difficulty of BRL items since only one object is the stimulus for each active-sort item. Experiment II, however, investigated the BRL passive sort; in this part of the test *S* is shown 12 groups of objects and asked, for each group, why the objects belong together. Since the passive-sort items consist of groups of objects, it is possible to obtain measures of mutual association for them.

The association measure used in this experiment was developed in a quite different area of investigation. The Index of Total Association (ITA), developed by Marshall and Cofer

(1963) for studying verbal recall, was adapted here for use in studying abstraction in the BRL Test. The rationale for use of the ITA in investigations of abstraction was the hypothesis that similar associative processes are essential variables in both abstraction and recall (Willner & Reitz, 1965).

The specific prediction of this experiment is that passive-sort items of the BRL Object Sorting Test having high ITA scores will be less difficult than those having lower ITA scores. This follows from the hypothesis that association functions as an essential variable in abstraction by eliciting the associative neighborhood containing the required concept. (Since groups of objects with a higher ITA score share more mutual associations, they are assumed to be more closely clustered in the same associative neighborhood than those groups having a lower ITA score.)

Method

The stimulus materials used were the 33 objects of the BRL Test. The *S*s, 60 introductory psychology students from Washburn University, were tested in two groups of approximately equal size. All *S*s were asked to give three word associations to each of the 33 objects.

Results

The definition of ITA as used in verbal recall (Marshall & Cofer, 1963) may be paraphrased as follows when it is used with an object sorting test. ITA is the mean extent to which objects in an object sorting test-item elicit each other as direct associates and the extent to which they also elicit other associative responses in common.

Scoring

ITA values were obtained for 11 of the 12 BRL passive-sort items, excluding Item 6. (This item was excluded because the objects are arranged for *S* in a manner different from the other 11 items, i.e., they are set down for *S* in sequence in pairs.) The ITA was computed in a manner analogous to Marshall and Cofer's (1963, p. 416) method by adding the number of associations in common to two or more objects in the BRL item, plus the number of instances in which one object in the item directly elicited another object in the item, and dividing by the total number of associations elicited by the objects in the item. Since the ITA was constructed for situations involving only one association to each list member, and *S*s in this experiment gave three associations to each object, it might be expected that the ITA would be markedly inflated. This was not the case, however, since the actual result was only to multiply both numerator and denominator by three, thus cancelling the effect. One change was introduced into the computation of the ITA. The BRL includes three pairs of identical objects: two nails, two corks, and two sugar cubes. (Identical words are not included in the word lists with which the ITA is used.) Since the authors felt that the presence of a pair of identical objects in an item would unduly inflate the ITA, one member of each identical pair of objects was eliminated from BRL passive-sort items containing such pairs, for purposes of computing the ITA. The object in the identical pair which had the greatest number of associations was retained (*S*s sometimes did not give all the associations required for all objects).

Table 5 indicates the ITA for each of the 11 BRL items used (the PCS values are discussed in Object Sorting Experiment III below). Six of the items have an ITA greater than .500, which means that more than half the total association to these items was accounted for by mutual associations among the objects. These six items were considered high

TABLE 5

INDEX OF TOTAL ASSOCIATION (ITA) SCORES AND THE PROPORTION OF THE CONCEPT RECALLED IN SEQUENCE (PCS) FOR 11 BRL PASSIVE-SORT CONCEPTS

BRL item no.	Concept	ITA	PCS
10	Silverware	.851	.899
4	Tools	.737	.677
2	Metal	.696	.622
11	Toys	.688	.380
9	Smoking	.663	.670
5	Paper	.507	.576
3	Round	.414	.391
12	Square	.342	.417
1	Red	.324	.200
7	White	.271	.464
8	Rubber	.204	.130

Note.—Item 6 has been omitted (see text).

ITA concepts, and the remaining five items were designated low ITA concepts.

The prediction that high ITA BRL items are less difficult than low ITA items was investigated by referring to the raw data used by Rapaport (1945), the scores of 54 Kansas State Highway Patrol officers. Rapaport's *S*s' scores for the 11 BRL passive-sort items investigated here were divided into two groups: a 6-item high ITA group and a 5-item low ITA group. Mean scores for the number of high and low ITA items correct were used, so that the unequal numbers of items would not bias the results. The mean number of high and low ITA items correct were, respectively, .927 and .663 ($p < .001$, t test), which means that 93% of high, but only 63% of low, ITA items on the average were solved correctly. Nine *S*s solved all 11 items correctly; of the remaining 45 *S*s the difference was in the predicted direction for all but 2.

CONCEPTUAL ORGANIZATION OF RECALL

A second aim of this research is empirical investigation of the important clinical hypothesis of the "conceptual organization of memories" (Rapaport, 1951). Several recent experiments in verbal recall imply that the organization of recall is affected by conceptual relations among the material to be recalled. Thus, when Bousfield (1953) obtained lists including categorized subgroups of words (e.g., animals, cities, etc.), arranged the words in random order, and asked *S*s to recall the lists, he found that in recall *S*s reassembled the words into the original categories.

Cohen, Bousfield, and Whitmarsh (1957), and Bousfield, Cohen, and Whitmarsh (1958) showed that the degree of conceptual reorganization of recall depended largely on the associative meaning of the words used. Cofer (1964) offered the proposition that when

> experimenter provided associational and categorical relations between members of a word pair provide a basis for clustering in free recall alternative to the bases, associational or otherwise, the subject will use to effect subjective organization or idiosyncratic pairing [p. 21].

The experimental techniques pioneered by the authors cited offer the possibility of precise empirical investigations of the conceptual organization of memory.

Willner and Reitz (1965) investigated the hypothesis that associative processes function as a common mediating process for abstraction and verbal recall, and also suggested further investigation of the conceptual organization of recall. Deese's (1959) experiment demonstrating the strong effect of association upon verbal recall was partially replicated with abstraction substituted as the dependent variable. Deese had demonstrated that lists of words having high mutual association were better recalled than lists with "zero" mutual association. Willner and Reitz used Deese's word lists as the stimuli in a word-sorting task in which *S*s were instructed to sort the words in each list into as many or as few groups as they wished, and to give a reason for each group. They predicted that (*a*) Deese's word lists having high mutual association would be sorted into fewer groups than those having "zero" mutual association; (*b*) the number of groups formed would be negatively correlated with a measure of mutual association; and (*c*) the number of groups formed would be negatively correlated with the number of words recalled for the lists. The confirmation of all three predictions strongly supported the hypothesis and is also in accordance with the hypothesis that stimuli which are more closely clustered in an associative neighborhood (i.e., have higher mutual association) will be less difficult both to recall and to group conceptually.

OBJECT SORTING EXPERIMENT III

This experiment extended the investigation of Willner and Reitz (1965) in three ways: (*a*) The stimuli used were objects rather than words; (*b*) The substitution was made in the opposite direction from Willner and Reitz (1965), that is, an abstraction task was transformed into a test of recall in Experiment III, rather than a recall task transformed into an abstraction test as in the earlier study; (*c*) The conceptual organization of recall was investigated by asking *S*s to recall the objects from the BRL Object Sorting Test after brief exposure, and ascertaining the extent to which their recall was conceptually organized,

Method

The stimulus materials were the 33 objects from the BRL Object Sorting Test. The *S*s, 25 psychology students from Washburn University, were group tested in two groups of approximately equal size. They were told that they would be shown a large number of objects for several minutes and then asked to list all the objects they remembered in the order they remembered them. They were then shown the 33 objects, arranged on a table so that no two closely adjacent objects were members of the same conceptual group. They were allowed to look at the objects for 4 minutes, after which the objects were covered up and each *S* was allowed 5 minutes to list the objects he remembered on a sheet of paper alongside numbers from 1 to 33.

Predictions

1. Recall of the BRL objects will be conceptually organized, that is, *S*s will group the objects in recall in accordance with the conceptual categories of the BRL passive sort.
2. *S*s will use high ITA concepts more often than low ITA concepts to organize recall of the objects. A greater proportion of the objects should thus be recalled in sequence in high ITA concepts than in low ITA concepts.
3. High ITA concepts will be used earlier than low ITA concepts in organizing *S*s' recall of the objects.

Results

Scoring—Prediction 1. The aim here was to measure the conceptual organization of *S*'s recall by determining the number of conceptually related sequences (CRSs) of objects he recalled. A CRS was defined as the maximum number of objects recalled in a conceptually unbroken sequence, that is, adjacent objects within the sequence had to be conceptually related. (There was no requirement that all objects within the CRS be instances of the *same* concept, only that adjacent objects share a common concept.) Objects were considered conceptually related only if, (*a*) they were members of the same BRL passive-sort conceptual group, or (*b*) two identical objects were recalled in sequence. (There are three such pairs of identical objects in the BRL: two nails, two sugar cubes, and two corks.) The degree of conceptual organization of *S*s' recall is inversely related to the number of CRSs for any given number of objects recalled, since the larger the number of CRSs the smaller is the number of objects included in each CRS.

Prediction 1, that *S*s' recall will be conceptually organized, may be stated more specifically, that is, the objects recalled by each *S* will be organized into fewer CRSs than would the same objects if recalled in random order. This prediction was tested by rearranging the objects recalled by each *S* into a random sequence using a table of random numbers (10 such random sequences were arranged for each *S*), and ascertaining the number of CRSs for the mean of the 10 random arrangements and also for the actual recall. The mean number of CRSs for the actual recall and random organization conditions was 8.52 and 18.91, respectively ($p < .001$, *t* test). The number of objects in the largest CRS group was also ascertained for both the actual recall condition and the mean of 10 random arrangements. The mean number of objects for the former was 9.08 and for the latter 3.91 ($p < .001$, *t* test). The strength of the effect is suggested by the observation that the results for both the number of CRSs, and the number of objects in the largest CRS were in the predicted direction for all 25 *S*s.

Scoring—Prediction 2. The aim here was to measure the proportion of objects in high and in low ITA concepts which were recalled in sequence. The first step in obtaining this score was to count the number of objects recalled in a sequence which included a common concept. A sequence of objects often includes more than one concept, and each concept common to at least two adjacent objects was scored (with two exceptions noted below). Thus if *S* recalled in sequence: a white rectangular filing card, a white square piece of cardboard, a red paper circle, a red rubber ball, a red round rubber sinkstopper, and a round bell, he would be counted as including two of the six white objects, two of the six rectangular objects, three of the five red objects, four of the six round objects, and two of the four rubber objects. The two exceptions referred to above are, first, when only two identical objects were grouped together (e.g., the two nails) they were not scored as an instance of a common concept (e.g., metal) because it seemed very likely that they were recalled as *two nails* rather than *two metal* *objects*. The second exception is that when a concept was split into two or more separated sequences, for example, the real and toy silverware recalled separately, the number of objects scored for any sequences after the first one was reduced by one. The rationale for this procedure was that a concept split into two or more separate sequences should obtain a lower score than the same concept preserved intact in one sequence. The final score used for each concept was obtained by dividing the number of objects recalled in sequence for each BRL concept by the total number of objects included in that concept. This score is the proportion of objects included in the concept, recalled in sequence (PCS).

BRL Item six, things grouped in pairs, was not investigated here since ITA scores were not obtained for it (for the reasons mentioned in the preceding experiment). The remaining 11 BRL passive-sort concepts were divided into high and low ITA groups. Since there were six high ITA concepts and five low ITA concepts, mean PCS scores for each group were used so that the unequal numbers of concepts would not bias the results. The mean PCS scores for high and for low ITA concepts, respectively, were 63.73 and 32.04 ($p < .001$, t test), which means that 64% of the objects in high ITA concepts, but only 32% of those in low ITA concepts on the average, were recalled in sequence. Moreover the difference was in the predicted direction for 24 of the 25 *S*s. Table 5 shows that the ITA and PCS scores for the 11 BRL passive-sort concepts were quite similar, with rank-difference correlation .77 ($p < .005$, one-tailed).

Scoring—Prediction 3. Here the intent was to ascertain the temporal order in which *S*s used the 11 concepts when recalling the objects. Since *S*s were asked to list all objects in order of recall, it was a simple task to note the concept used first, second, etc. For example, if the first five objects were recalled by *S* in the order: two nails and three large metal silverware, the concept *metal* was used first and *silverware* second. When the first objects were included in more than one concept, an average rank was given. For example, if *S* recalled the above objects in this order: three silverware and two nails, the concepts *metal* and *silverware* would be given a rank of 1.5 since the silverware are both metal and silverware.

Several additional rules were used. If the first instance of a concept was an isolated pair of identical objects (nails, corks, sugar cubes) with no other adjacent member of the same concept, then these identical objects were not scored as an instance of the concept. A second rule was that only the first instance of a concept used was scored, since the aim was to ascertain the earliest use of each concept. Concepts which were not used at all in *S*'s recall were scored as if they had been given last, with provision made for ties if more than one concept was not used.

The 11 BRL passive-sort concepts were divided into high and low ITA groups, as above. The score used was the rank for temporal order of *S*'s use of the concept (with the first concept used being given a rank of 1, etc.). Since there were six high ITA and only five low ITA concepts, mean ranks were used for each group so that unequal numbers of concepts would not bias the results. The mean ranks for high and low ITA concepts, respectively, were 4.62 and 7.66 ($p < .001$, t test). The difference was in the predicted direction for 22 of the 25 *S*s. The association between ITA and temporal order of use of the concepts in recall is also indicated by the rank-difference correlation of .88 ($p < .005$, one-tailed) between the two variables.

DISCUSSION

Similarities tests have been criticized for being too vulnerable to solution by simple association processes, and thus insensitive to impairment and even deterioration of abstract reasoning in patients known to have such impairment. Rapaport (1945) described the Wechsler-Bellevue Similarities Test as solvable by automatic response processes, by "the empty shells of concepts [p. 395]." Wechsler himself (1958) commented about the WAIS Similarities Test:

> Some of the responses awarded maximum credit turn out in many instances to be merely verbal associations—for example, Banana—Apple, "Both fruit"; or even Fly—Tree, "Both living things" [p. 182].

This problem in accurate assessment of abstraction seems largely attributable to the effects of S-C association, which may too automatically produce the required concept. Tests in which mutual association is the major association variable might well be more discriminating measures of abstraction because the association link, rather than eliciting the required concept, would elicit the associative neighborhood common to both stimuli. Similarity, in a similarities test, could then be operationally defined as the degree of overlap of the associative networks evoked by each word, and would be estimated by the degree of mutual association between the two words. Such an operational definition of similarity would be consonant with recent research on the structure of associative meaning, for example, Deese (1962). Thus, two major differences between S-C and mutual association are that only the former directly produces the correct answer to the item, whereas only the latter deals with actual similarity (associative similarity) between the stimulus words.

The confirmation of all predictions in the five experiments reported strongly supports the major hypothesis, that association functions in abstraction by eliciting the required concept directly (S-C association), and also by eliciting the associative neighborhood containing the required concept (mutual association). The results also demonstrate that the relation between association and abstraction test responses applies to a less verbal test, the Object Sorting Test, as well as to a more verbal one, the Similarities Test.

In addition, the confirmation of all predictions in Similarities Experiments I and II supports the hypothesis that paired presentation of the stimulus words permits a focusing effect for associations to both words. The magnitude of the increase in "correct scores" with paired presentation of words is indicated by the mean scores for single- and paired-word presentation, respectively, 5.11 and 10.72. The equivalent WAIS scaled score for the mean for paired presentation (10.72 rounded off to 11) is 9. Since the mean scaled score for the average person on the WAIS Similarities Test is 10, the score of 9 obtained by S-C association under paired presentation is almost as high as the score of the average person. (Since the Similarities Test words are presented to *S* in pairs, paired-presentation association scores seem the more appropriate ones to use.) Several other factors should be considered in evaluating the meaning of this result. First, the score used was the total score for three associations given by *S*, so that each *S* had three chances to obtain a "correct" answer. Although the maximum score for a WAIS Similarities Test item is two, it was possible for *S*s in this association condition to obtain a score of four since all "correct" scores were summed. (Several *S*s actually accomplished this by giving as their three associations, one good—2-point—answer and two mediocre—1-point—answers.) This method thus permits higher scores than does the Similarities Test. Secondly it is also possible that some *S*s among the sample of college students may have developed a set for associating in terms of similarities, even though they were asked to give only the first three words that came to mind. Yet, even when these factors are considered, it is clear that *S*s may obtain quite high scores by S-C association.

These results have obvious implications for constructing more clinically sensitive similarities tests, that is, reducing or, if possible, eliminating the effect of S-C association in the test. Experiments are currently under way which aim at developing new similarities tests free of S-C association and having mutual association as the major association variable. Mutual association is stressed because of two hypotheses: 1. Association is an essential variable in abstraction, and 2. The degree of shared-associative meaning (mutual association) can serve as an operational definition of similarity in similarities tests.

The confirmation of the predictions in Object Sorting Experiments I and II strongly supports the major hypothesis in the case of the BRL Object Sorting Test. The results of Experiment II also make it very clear that ITA is an effective predictor of item difficulty in an object sorting test. In fact, the classification of the 11 items into high and low ITA concepts is very similar to Rapaport's empirical classification of the items as easy and difficult. (Five of the six high ITA items are classified by Rapaport as easy and four of the five low ITA items are classified by him as difficult.)

These results emphasize the major role

played by association in both abstraction and recall. This is especially impressive when one considers that the ITA, which was developed as a predictor of verbal recall based on verbal association data, is used in this experiment as a predictor of difficulty of object sorting test items based upon object association data, and that despite this great change it has functioned very efficiently.

Object Sorting Experiment III offers the possibility of investigating the conceptual organization of recall by using the objects from a clinical test of abstract reasoning (the BRL), but presenting them to *S* as a memory task. Experiment III thus differs from the usual abstraction task, in which the stimuli are typically present and available to *S* when he takes the test, since in Experiment III the stimuli are removed from him and he does not even know that abstraction is being tested. Since the BRL test specifies the conceptual relationships among the objects, one may investigate whether *S*'s recall is conceptually organized, that is, in terms of the concepts measured by the test.

Experiment III investigated two major hypotheses. The first was that *S*'s recall would be conceptually organized. The second hypothesis dealt with prediction of which concepts *S* would use to organize his memories and also the temporal order in which the concepts would be used. The second hypothesis, more specifically, was that the ITA value for the objects included in the concept would determine: (*a*) which concepts were used to group the objects in recall, and (*b*) the temporal order in which the concepts would be used in recall. The confirmation of all three predictions provided strong support for the hypothesis of conceptual organization of recall and for the hypothesis that association (ITA) is an essential variable in determining which concepts will be used to organize recall. The strength of these effects is suggested by the very large proportion of *S*s for whom the three predictions were confirmed and by the high rank-order correlations (.77 and .88, respectively) between ITA and the proportion of the concept recalled in sequence (PCS), and between ITA and the temporal order of use of the concepts in recall. Table 5 shows the striking similarity between ITA and PCS scores for each concept.

The object-recall procedure used in this experiment may also have in itself considerable potential as a clinical test of abstraction. The authors have used the object-recall procedure as a third part of the BRL Object Sorting Test, administered before the active and passive sort, and found that it provides useful clinical information. For example, some patients who use very narrow concepts in the two parts of the BRL (e.g., one patient who grouped only the small fork with the large fork and excluded the remaining silverware) behave quite differently in the object-recall condition. The patient just cited listed all six silverware in sequence in the object-recall condition. The explanation of this apparent paradox was that he was obsessively hypercritical when he knew his abstract reasoning was being assessed, but since he was not aware that his abstraction was being investigated in the object-recall condition he performed more freely. It is quite clear at this point that the object-recall procedure is a much more subtle test of abstraction than are the two traditional parts of the test. The authors have the definite impression that research such as this may also provide many other useful clinical tools.

This research illustrates the likelihood that investigations of associative meaning may lead to a mutual enrichment of clinical and experimental psychology. The contribution to clinical psychology is clear, both in terms of development of precise methods for investigating major subject areas until now refractory to such investigation, and also in terms of the potential development of new clinical instruments which might add to the clinician's armamentarium.

The contribution to experimental psychology could reside in the area of interrelations between abstraction and recall, in what might be termed the conceptual organization of recall. This interrelation between abstraction and recall is illustrated by two experiments, Willner and Reitz (1965) and Object Sorting Experiment III in this study. In the former, stimulus materials from an experiment in verbal recall were substituted as stimuli in an experiment in verbal abstraction, and by means of a knowledge of the associative meaning of the stimuli, precise predictions were confirmed about their functioning as an abstraction task. Experiment III performed the substitution in the reverse direction, that is, stimuli from an abstraction task were used as a recall task, and moreover the stimuli were objects rather than words. The very strong confirmation of all predictions in this experiment underscores the importance of the overlapping areas of abstraction and recall.

REFERENCES

BOLLES, M., ROSEN, G., & LANDIS, C. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia. *Psychiatric Quarterly*, 1938, **12**, 733–737.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychological Reports*, 1958, **4**, 39–44.

COFER, C. N. On some factors in the organizational characteristics of free recall. Technical Report No. 1, Contract Nonr 656 (30), Pennsylvania State University and the Office of Naval Research, 1964.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, Contract Nonr 631 (00), Office of Naval Research and University of Connecticut, 1957.

COLEMAN, E. B. Verbal concept learning as a function of instructions and dominance level. *Journal of Experimental Psychology*, 1964, **68**, 213–214.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, **5**, 305–312.

DEESE, J. On the structure of associative meaning. *Psychological Review*, 1962, **69**, 161–175.

FREEDMAN, J. L., & MEDNICK, S. A. Ease of attainment of concepts as a function of response dominance variance. *Journal of Experimental Psychology*, 1958, **55**, 463–466.

MARSHALL, G. R., & COFER, C. N. Associative indices as measures of word relatedness: A summary and comparison of ten methods. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 408–421.

MEDNICK, S. A., & HALPERN, S. Ease of concept attainment as a function of associative rank. *Journal of Experimental Psychology*, 1962, **64**, 628–630.

POLLIO, H. R. Composition of associative clusters. *Journal of Experimental Psychology*, 1964, **67**, 199–208.

RAPAPORT, D. *Diagnostic psychological testing*. Chicago: Year Book Medical, 1945.

RAPAPORT, D. *Organization and pathology of thought*. New York: Columbia University Press, 1951.

SCHEERER, M. Spheres of meaning: An analysis of stages from perception to abstract thinking. *Journal of Individual Psychology*, 1959, **15**, 50–61.

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *Journal of Experimental Psychology*, 1956, **51**, 229–238.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

WECHSLER, D. *The measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins, 1958.

WHORF, B. L. *Language, thought, and reality*. New York: Massachusetts Institute of Technology Press, and Wiley, 1956.

WILLNER, A. E., & REITZ, W. E. Association strength as a mediating variable in verbal categorizing and recall: Preliminary report. *Psychological Reports*, 1965, **16**, 310.

(Received October 5, 1965)

Beginning with the 1967 volume, this Journal will consider manuscripts for publication as monographs. Monograph Supplements will appear under separate covers, but will be mailed with issues of the Journal to all subscribers.

Monographic manuscripts will be judged by the same standards of quality as articles in the Journal. In addition, they are expected to present comprehensive experimental investigations, frequently reporting a series of experiments, or to provide more extensive theoretical integration of a topical area than is possible in the space available to articles.

Authors of monographs whose content is germane to the area of this Journal are invited to submit manuscripts to the editor. Because of space limitations, however, probably not more than one monograph per year can be accommodated.

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 328–334

URINARY EXCRETION OF ADRENALINE AND NORADRENALINE DURING PROLONGED PERCEPTUAL DEPRIVATION[1]

JOHN P. ZUBEK AND W. SCHUTTE

University of Manitoba

Ss who successfully completed a week of perceptual isolation showed no significant changes in urinary excretion of adrenaline and noradrenaline relative to a group of control Ss. They also showed no changes on a behavioral measure of subjective stress and mood. Isolation "quitters," i.e., those who failed to complete the prescribed period, also showed no change in noradrenaline. Their urinary excretion of adrenaline, however, increased, but only in Ss who terminated isolation late in the 1-wk period. Perhaps the most important characteristic of the isolation quitters was the presence of a significantly lower baseline level of adrenaline, relative to that of the successful experimental Ss, not only prior to isolation but also many months after its termination. No such difference was found for noradrenaline. These results seem to suggest that isolation quitters may be biochemically or "constitutionally" different from volunteers who can successfully complete a prolonged period of perceptual isolation.

Various investigators have reported that prolonged exposure of human subjects (*S*s) to either sensory or perceptual deprivation is a stressful experience (Bexton, Heron, & Scott, 1954; Myers, Murphy, & Terry, 1962; Smith & Lewty, 1959; Zuckerman, 1964). If isolation is as stressful as these reports suggest, one might expect an increase in the activity of the sympathetic adrenal system, particularly in the light of numerous studies indicating its involvement in various stressful situations of a psychological nature (e.g., Elmadjian, Hope, & Lamson, 1957; Frankenhaeuser & Patkai, 1964; Funkenstein, King, & Drolette, 1957; Levi, 1964; Selye, 1950; Stoner, 1960). Unfortunately, little relevant data are available on the effects of isolation. Only one exploratory study, performed at McGill University, has been concerned with possible changes in adrenal activity during prolonged periods of isolation (Murphy, Kurlents, Cleghorn, & Hebb, 1955). Urinary measures of 11-oxycorticoids were taken from nine *S*s before and during 1½–6 days of perceptual deprivation (unpatterned light and constant noise). Since no consistent increase in the excretion of corticoids occurred, it was concluded that under isolation the "adrenal cortex is not activated to a greater degree than it is by the minor exigencies of everyday life [p. 1063]." In view of these negative results, a study was conducted to determine whether the activity of the sympathetic adrenomedullary system, as measured by the urinary excretion of adrenaline and noradrenaline (catecholamines), would similarly be unaffected by prolonged periods of deprivation. Although catecholamine determinations have been performed in three earlier studies, all of them employed deprivation periods of less than a day (Cohen, Silverman, & Shmavonian, 1962; Mendelson, Kubzansky, Leiderman, Wexler, Dutoit, & Solomon, 1960; Schaefer, 1964). Furthermore, the results revealed no consistent pattern of biochemical changes during the isolation period.

METHOD

A group of 31 male university students were placed individually in an air-conditioned isolation chamber for a prescribed period of 1 week. Of this group, 18 successfully endured the week while 13 were unable to do so (*M* duration = 56 hours). This latter group will be referred to as the unsuccessful *S*s or "quitters." All *S*s wore translucent goggles and a set of earmuffs through which white noise was constantly presented somewhat above the threshold of hearing. Further details of the isolation chamber and the experimental procedure are available in an earlier publication (Zubek, Aftanas, Hasek, Sansom, Schludermann, Wilgosh, & Winocur, 1962). Since the recumbent position, which *Ss* maintained most of the time because they had nothing to do, is known to decrease the output of both adrenaline and noradrenaline (Euler, Luft, & Sundin, 1955; Sundin, 1958), a group of 20 recumbent controls were placed individually inside the chamber for 1 week. Two of them terminated the condition prematurely. These control *Ss* were permitted the use of reading material, a radio, and were frequently visited by the experimenters (*Es*). Furthermore, the walls were covered with brightly colored pictures. The two doors to the chamber were left open at all times. Both the controls and the experimental *Ss* were housed in the same chamber in order to control for the possible effects of suggestion resulting from their placement in a novel or unusual laboratory situation. All of the controls had initially volunteered as experimental *Ss*.

---

[1] This research was supported by the Defence Research Board of Canada (Project 9425-08) and by Grant MHO8748 from the National Institute of Mental Health, United States Public Health Service. The authors wish to express their appreciation to M. Nickerson, Department of Pharmacology, University of Manitoba Medical School, for providing facilities for the biochemical determinations.

Urine samples were taken at 24-hour intervals for 3 days prior to the experiment, at daily intervals during the experiment, and for 2 days after its termination. Thus, twelve 24-hour urine samples were taken from each experimental and control *S* who successfully completed the 1-week period. A further 24-hour urine sample was obtained from 18 recumbent controls and 14 of the 31 experimental *Ss* approximately 6 months after completion of isolation (8 successful *Ss* and 6 quitters). The remainder could not be tested since they were no longer in the city. The catecholamine determinations were performed by the fluorimetric procedure of Euler and Lishajko (1961) and were expressed as total excretion per 24-hour period.

In addition to these chemical analyses, an appraisal was made of *Ss*' affective state by two paper and pencil tests. The first was the Subjective Stress Scale (SSS) which has been shown to differentiate well between individuals who have experienced situations inducing various degrees of stress (Berkun, Bialek, Kern, & Yogi, 1962; Myers, 1964). The SSS is a 15-item Thurstone scale, with higher scores indicating more adverse reactions. The second test, developed by Myers, Murphy, Smith, and Goffard[2] specifically for the isolation condition, is a measure of overall mood. It consists of a 114-item adjective check list which *S* answers as applying "not at all," "somewhat or slightly," or "mostly or generally" to him during a particular time period. Both the mood test and the SSS were administered at daily intervals before, during, and after the experimental period, that is, at intervals corresponding to the 24-hour urine samples.

RESULTS AND DISCUSSION

Figure 1 summarizes the results on noradrenaline, the peripheral neurohormone of the sympathetic nervous system (Euler, 1956). It can be seen that both the successful and unsuccessful experimental *Ss* (the quitters) show a noradrenaline pattern similar to that of the controls. An analysis of variance revealed that all three groups showed a significant decrease in mean urinary excretion of noradrenaline during the experimental period relative to their "pre-and post-" experimental levels (p's $< .001$). None of the differences among the three groups, however, were statistically significant.

Figure 2 summarizes the results on adrenaline which is the main constituent of catecholamines secreted by the adrenal medulla. Again, there was no significant difference in mean excretion level between the successful experimental *Ss* and the controls relative to their pre- and postexperimental levels. Both groups showed a significant decrease in adrenaline while in the chamber (p's $< .02$). This decrease in adrenaline, together with that for noradrenaline as indicated in Figure 1, probably resulted from the recumbent position maintained most of the time by all *Ss* (Euler et al., 1955; Sundin, 1958). The quitters, on the other hand, were characterized by a complex pattern of changes. First, the adrenaline level during isolation was related to their endurance. The early quitters, that is, those who terminated the condition within the first 2 days, showed an adrenaline excretion level which was not significantly different from either their preisolation (Preday 2 and 3) or postisolation (Postday 2) baseline levels. The late quitters, however, showed a significant increase in adrenaline relative to their "pre- and post-" baseline level after the second day of isolation (p's $< .05$) but not during the first 2 days. A possible explanation of this difference between the two types of *Ss* may be that the early quitters terminated isolation at the first indication of discomfort and unpleasantness, and thus showed no change in adrenaline level. On the other hand, the adrenaline increase in the late quitters may be a reflection of an accumulating, growing

---

[2] T. I. Myers, D. B. Murphy, S. Smith, & S. J. Goffard. Experimental Studies of Sensory Deprivation and Social Isolation. Unpublished Manuscript.

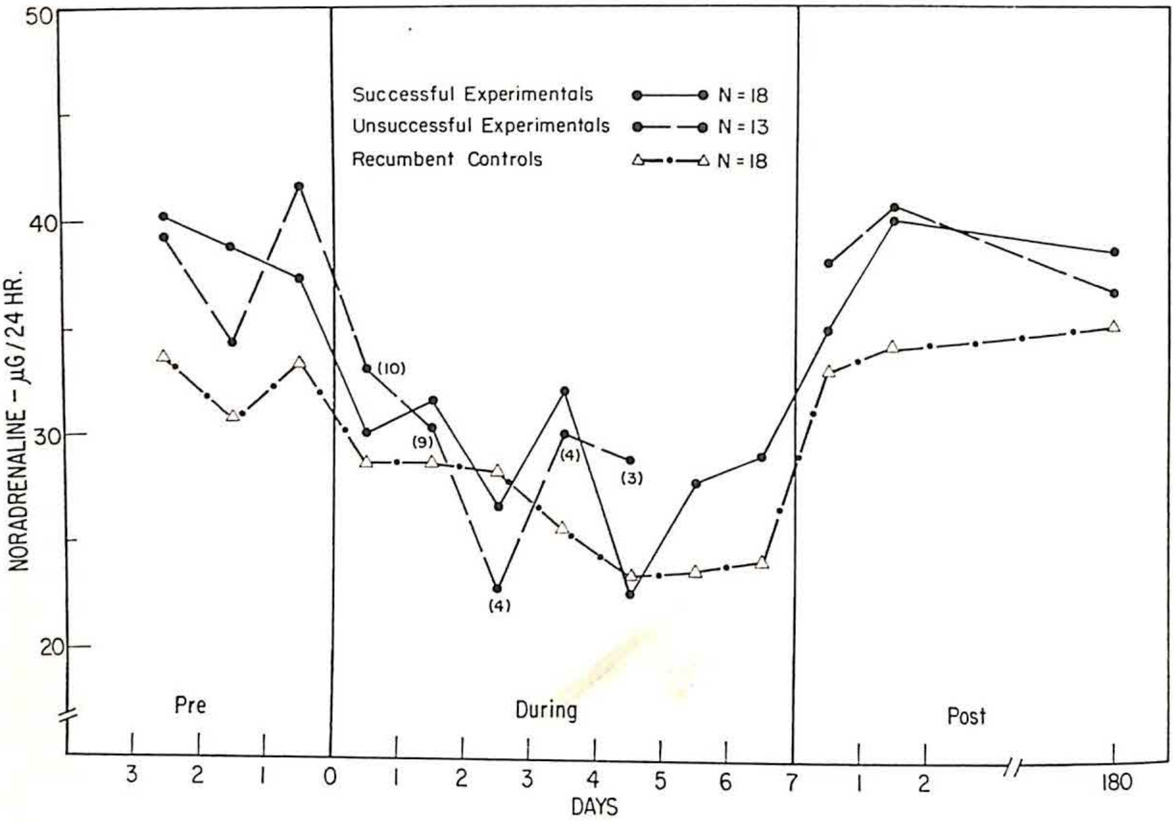


FIG. 1. Urinary excretion of noradrenaline (24-hour output) in two experimental groups and one control group before, during, and after a 1-week period. (The figures in parentheses refer to the number of unsuccessful experimental subjects or "quitters" of the original group of 13 who remained in isolation for a specified number of days, for example, 10 subjects completed at least 1 day; the remaining 3 terminated isolation very early in the first day, voiding either no urine or an insufficient amount to permit a biochemical analysis. Furthermore, of the 10 subjects only 3 remained in isolation to the end of the fifth day.)

emotional tension developing over time. Furthermore, this adrenaline increase may have produced certain psychological changes which precipitated the termination. This interpretation is supported to some degree by the fact that the scores on the SSS were significantly higher in the late than in the early quitters ($p < .01$).

A second characteristic of the unsuccessful *S*s, as indicated in Figure 2, was the presence of a significant increase in adrenaline on the day *prior* to isolation, relative to Preisolation Days 2 and 3, and relative to the changes shown by the other two groups of *S*s (p's $<$ .05). A similar increase in preisolation noradrenaline level, although not as noticeable, can be seen in Figure 1. Further analysis of this phenomenon revealed that the quitters who showed the greatest preisolation increase in adrenaline tended to terminate isolation the earliest. The correlation, however, was not significant (rho $= -0.28$). The presence of a high catecholamine level prior to an experiment has also been noted in other studies, for example, immediately before centrifuge (Goodall & Berman, 1960) or drug studies (Elmadjian et al., 1957), and is believed to be indicative of apprehension about experimental participation. This "apprehension" effect was clearly evident in the results of the stress scale (see Figure 3). However, contrary to what might be expected from the catecholamine data, both the isolation quitters and the successful experimentals showed a significant increase in stress score on the day prior to the experiment and relative to the preceding 2 days (p's $< .05$).

The third, and perhaps the most important, characteristic of the quitters was the possession of a preisolation baseline level of adrenaline (Preday 2 and 3) which was significantly lower relative to that of the successful experimental *S*s ($p < .01$). This low excretion level was also seen 2 days after termination of isolation ($p < .01$) as well as 6 months later ($p < .05$). No such difference between the two types of experimental *S*s occurred in the noradrenaline baseline level. Since this adrenaline baseline difference seems to be a stable characteristic, persisting over many months, it would appear that isolation quitters may be biochemically or "constitutionally" different from volunteers who can endure a prescribed period of prolonged isolation. In view of these results, further research is required to determine whether other biochemical differences may exist between these two types of isolation *S*s. The presence of this biochemical difference suggests that these *S*s may also differ in personality. This does not appear to be the case. Administration of the Rorschach, MMPI, Edwards Personal Preference, and the Thurstone Temperament Schedule prior to isolation indicated that no single personality characteristic measured by these tests significantly differentiated the successful from the unsuccessful *S*s, a finding also observed in two earlier studies from this laboratory (Hull & Zubek, 1962; Wright & Zubek, 1966).

The adrenaline excretion of the recumbent controls is interesting in two respects: First, their pre- and postbaseline levels fell in the range between the baseline levels of the successful and unsuccessful isolation *S*s. This might be expected since these controls can be considered to consist of both potential isolation quitters and successful *S*s, particularly since they all volunteered initially as experimental *S*s. Second, both *S*s who terminated the control condition prematurely showed an abnormally low baseline level of adrenaline before and after the experiment (range = 0.50–2.70 μg. per 24 hours). Al-

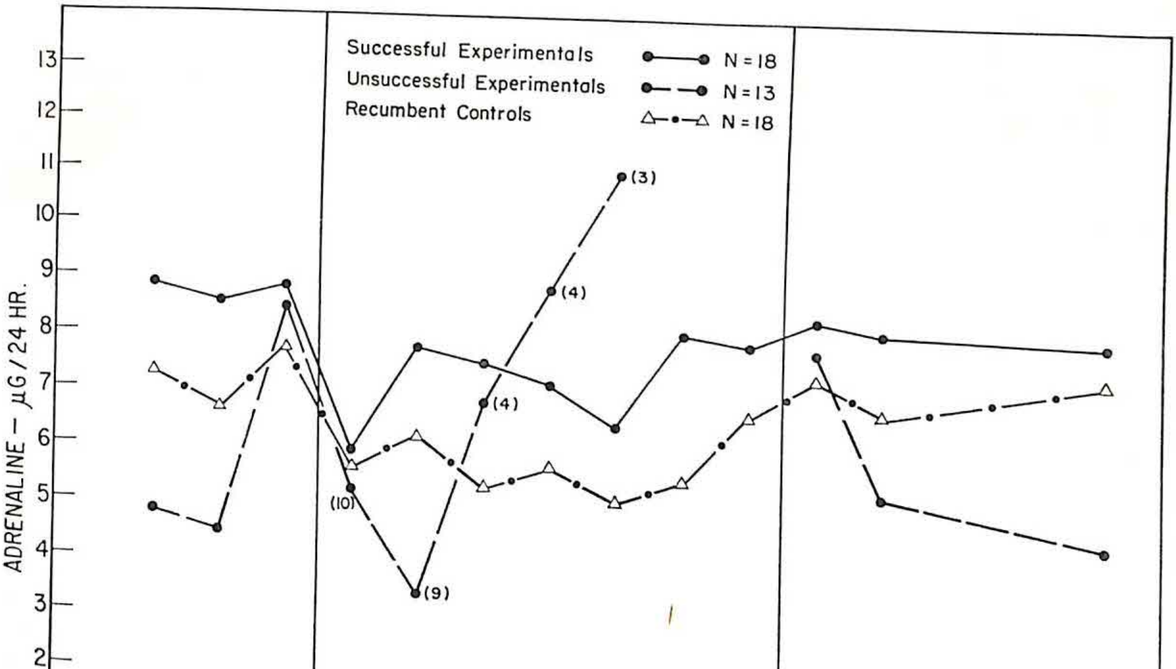


FIG. 2. Urinary excretion of adrenaline (24-hour output) in two experimental groups and one control group before, during, and after a 1-week period.

though these results are based on only two "control quitters" they do suggest that low adrenaline baseline levels may be predictive not only of isolation intolerance but also of the inability to successfully endure other types of related conditions, for example, prolonged bed rest or restricted motor activity.

An analysis of the 24-hour urine volume of the successful experimentals revealed a progressive decrease in excretion during the week of isolation, a decrease significantly greater than that shown by the recumbent controls ($p < .001$). This decreased urine volume was still present 2 days after termination of isolation ($p < .01$) but was back to the preexperimental level 6 months later. No significant difference in total fluid intake of the two groups of *S*s occurred during the 1-week period. A similar progressive decline in urinary volume, independent of fluid intake, has also been reported by Winters (1963) in monkeys who were placed in darkness and silence for 14 days. Since water consumption was not reduced in either experiment, this decreased urinary volume may be a primary response possibly "triggered" by a greater production of the antidiuretic hormone of the pituitary gland. This suggestion of a possible pituitary disturbance is further supported by a recent study in which an increased production of the pituitary thyroid-stimulating hormone occurred after 8 hours of sensory deprivation (Zuckerman, Persky, Hopkins, Murtaugh, Basu, & Schilling, 1966).

A decrease in urine volume was also observed in the unsuccessful experimental *S*s with the exception of the four late quitters who showed a considerable increase after the second day of isolation. No significant correlations were found between urine volume and adrenaline and noradrenaline excretion for the preexperimental, experimental, and postexperimental periods, a finding also reported in an earlier study on short-term perceptual deprivation (Mendelson et al., 1960).

In Figure 3 it can be seen that the successful isolation *S*s show a higher subjective stress score, relative to the recumbent controls, during the 1-week experimental period. This difference, however, was not significant with respect to either their pre- or postexperimental baseline scores (Preday *2* and 3; Postday 2). On the other hand, the isolation quitters show a significant increase in stress score after the first day of isolation, relative to the other groups of *S*s (p's $< .01$). These results, therefore, closely parallel those of the catecholamine determinations. Figure 3 also shows a marked "apprehension" effect, as indicated by significantly higher stress scores, in the isolation but not in the control *S*s on the day prior to the experiment and relative to the *2* preceding days (p's $< .05$). This effect is shown by both the isolation quitters and the successful *S*s and to approximately the same degree.

The results on the mood scale indicated that both the successful and unsuccessful experimental *S*s were characterized by a more negative mood during isolation than were the controls. None of the differences, however, were significant. Furthermore, the mood scale revealed no indication of a preexperimental apprehension effect in any group of *S*s.

These findings, together with the McGill data on 11-oxycorticoids, suggest that neither the adrenocortical nor the sympathetic adrenomedullary systems are activated to any noticeable degree by prolonged periods of perceptual deprivation. Although isolation *S*s may report a variety of affective disturbances, as indicated in some of the literature, these do not appear to be reflected either on the behavioral measures or in the biochemical indexes of stress that have so far been employed. It is possible, though, that other chemical substances such as 17-ketosteroids and 17-ketogenic steroids may increase during prolonged isolation, especially since the excretion of both these adrenocortical substances was recently shown to be greater during short-term sensory deprivation (Zuckerman et al., 1966). It is also possible that increased activity of the sympathetic adrenal system might have occurred in the successful *S*s if the isolation condition had been made unusually severe by elimination of *all* intercommunication between the isolated *S* and *E*s. Some indirect evidence for this hypothesis has been provided by the Russian investigators, Gorbov, Miasnikov, and Yazdovsky (1963), who reported an increase in 17-ketosteroids in *S*s socially isolated for 15 days

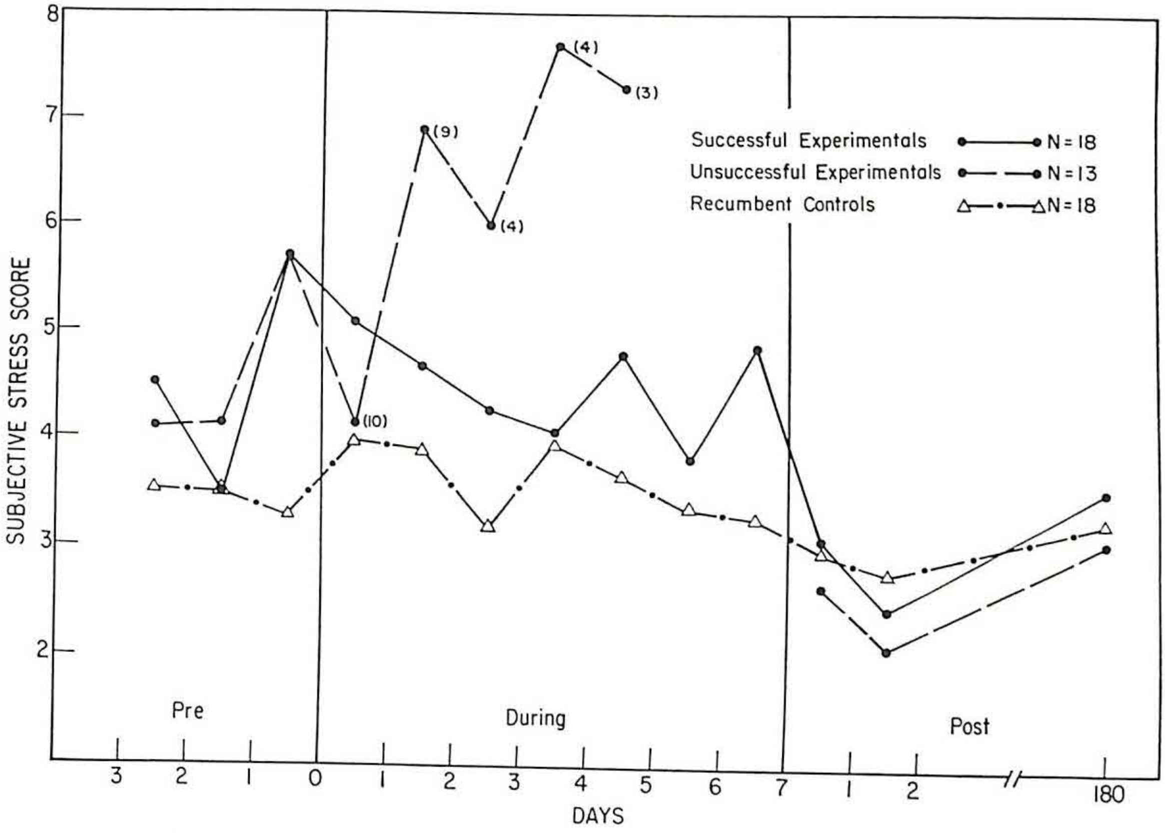


FIG. 3. Subjective stress scores of two experimental groups and one control group at daily intervals before, during, and after a 1-week period. (The higher scores on the ordinate indicate greater subjective stress.)

under a condition of no intercommunication. However, in experiments where some degree of two-way communication was introduced, no changes in 17-ketosteroid output were observed. Severity of the perceptual deprivation condition may, therefore, be an important variable in accounting for these results.

Although *Ss* who successfully endured a week of perceptual deprivation showed no increase in urinary catecholamines, a heightened level was observed in the late isolation quitters. The results, however, were differential in nature. A significant increase in adrenaline but not in noradrenaline was found to occur. A similar differential pattern has also been observed in other "stress" conditions, for example, during water immersion (Goodall, McCally, & Graveline, 1964), performance of insolvable tests (Frankenhaeuser & Patkai, 1964), and in psychiatric patients during a staff interview (Elmadjian et al., 1957). Although the reasons for this differential pattern are not clear, such results might be expected in view of the different site of production of adrenaline and noradrenaline (Euler, 1956).

REFERENCES

BERKUN, M. M., BIALEK, H. M., KERN, R. P., & YOGI, K. Experimental studies of psychological stress in men. *Psychological Monographs,* 1962, 76(15, Whole No. 534).

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology,* 1954, 8, 70–76.

COHEN, S. I., SILVERMAN, A. J., & SHMAVONIAN, B. M. Psychophysiological studies in altered sensory environments. *Journal of Psychosomatic Research,* 1962, 6, 259–281.

ELMADJIAN, F., HOPE, J. M., & LAMSON, E. T. Excretion of epinephrine and norepinephrine in various emotional states. *Journal of Clinical Endocrinology,* 1957, 17, 608–620.

EULER, U. S. v. *Noradrenaline.* Springfield, Ill.: Charles C Thomas, 1956.

EULER, U. S. v., & LISHAJKO, F. Improved technique for the fluorimetric estimation of catecholamines. *Acta Physiologica Scandinavia,* 1961, **51,** 348–355.

EULER, U. S. v., LUFT, R., & SUNDIN, T. The urinary excretion of noradrenaline and adrenaline in healthy subjects during recumbency and standing. *Acta Physiologica Scandinavia,* 1955, **34,** 169–174.

FRANKENHAEUSER, M., & PATKAI, P. Interindividual differences in catecholamine excretion during stress. *Report of the Psychological Laboratory,* University of Stockholm, 1964, No. 178.

FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. E. *Mastery of stress.* Cambridge: Harvard University Press, 1957.

GOODALL, M., & BERMAN, M. L. Urinary output of adrenaline, noradrenaline, and 3-methoxy-4-hydroxymandelic acid following centrifugation and anticipation of centrifugation. *Journal of Clinical Investigation,* 1960, **39,** 1533–1538.

GOODALL, M., McCALLY, M., & GRAVELINE, D. E. Urinary adrenaline and noradrenaline response to simulated weightless state. *American Journal of Physiology,* 1964, **206,** 431–436.

GORBOV, F. D., MIASNIKOV, V. I., & YAZDOVSKY, V. I. (Strain and fatigue under conditions of sensory deprivation). *Zhurnal Vysshei Nervnoi Deiatelnosti, Pavlov,* 1963, **13,** 585–592. (In Russian)

HULL, J., & ZUBEK, J. P. Personality characteristics of successful and unsuccessful sensory isolation subjects. *Perceptual & Motor Skills,* 1962, **14,** 231–240.

LEVI, L. The urinary output of adrenaline and noradrenaline during different experimentally induced pleasant and unpleasant states. A summary. *Journal of Psychosomatic Research,* 1964, **8,** 197–198.

MENDELSON, J., KUBZANSKY, P. E., LEIDERMAN, P. H., WEXLER, D., DUTOIT, C., & SOLOMON, P. Catechol amine excretion and behavior during sensory deprivation. *Archives of General Psychiatry,* 1960, **2,** 147–155.

MURPHY, C. W., KURLENTS, S., CLEGHORN, R. A., & HEBB, D. O. Absence of increased corticoid excretion with the stress of perceptual deprivation. *Canadian Journal of Biochemical Physiology,* 1955, **33,** 1062–1063.

MYERS, T. I. Studies of the stressfulness of sensory deprivation and confinement. Paper read at Eastern Psychological Association, Philadelphia, April 1964.

MYERS, T. I., MURPHY, D. B., & TERRY, D. F. The role of expectancy in subjects' responses to sustained sensory deprivation. Paper read at the American Psychological Association, St. Louis, September 1962.

SCHAEFER, K. E. Counteracting effects of training in geometrical construction on stress produced by maximal sensory isolation in water immersion. *Aerospace Medicine,* 1964, **35,** 279. (Abstract)

SELYE, H. *Stress.* Montreal: Acta Inc., 1950.

SMITH, S., & LEWTY, W. Perceptual isolation using a silent room. *Lancet,* 1959, **2,** 342–345.

STONER, H. B. (Ed.) *The biochemical response to injury.* Springfield, Ill.: Charles C Thomas, 1960.

SUNDIN, T. The effect of body posture on the urinary excretion of adrenaline and noradrenaline. *Acta Medica Scandinavia,* 1958, Suppl. **336,** 1–59.

WINTERS, W. D. Various hormone changes during simulated space stresses in the monkey. *Journal of Applied Physiology,* 1963, **18,** 1167–1170.

WRIGHT, N. A., & ZUBEK, J. P. Use of the multiple discriminant function in the prediction of perceptual deprivation tolerance. *Canadian Journal of Psychology,* 1966, **20,** 105–113.

ZUBEK, J. P., AFTANAS, M., HASEK, J., SANSOM, W., SCHLUDERMANN, E., WILGOSH, L., & WINOCUR, G. Intellectual and perceptual changes during prolonged perceptual deprivation: Low illumination and noise level. *Perceptual & Motor Skills,* 1962, **15,** 171–198.

ZUCKERMAN, M. Perceptual isolation as a stress situation. *Archives of General Psychiatry,* 1964, **11,** 255–276.

ZUCKERMAN, M., PERSKY, H., HOPKINS, T. R., MURTAUGH, T., BASU, G. K., & SCHILLING, M. A comparison of stress effects of confinement and social isolation with and without perceptual restrictions. *Archives of General Psychiatry,* 1966, **14,** 356–365.

(Early publication received March 14, 1966)

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 335-344

PSYCHOPHYSIOLOGICAL REACTIONS TO FILMS BY CHRONIC SCHIZOPHRENICS:

I. EFFECTS OF DRUG STATUS[1]

MICHAEL J. GOLDSTEIN
University of California, Los Angeles
CHARLES W. ACKER, JOSEPH T. CROCKETT, AND JOHN J. RIDDLE
Veterans Administration Center, Los Angeles

40 chronic schizophrenic patients had their psychophysiological responses to 1 of 2 tension-arousing movies recorded. Following assignment to drug (thiordizane) or placebo status for 1 wk. they were retested on a similar film. The results were that the drug Ss showed a decrease in skin resistance and heart rate but not finger-pulse volume over their responses to the 1st showing while the placebo group showed opposite trends. Interview data collected after the films indicated no correlation between the autonomic changes and plot comprehension. Pilot work on word-association test responses indicates that the drug group showed a decrease while the placebo group showed an increase in idiosyncratic responses.

Despite repeated demonstrations with acute (Psychopharmacology Study Center Group, 1964) and chronic (Casey, Bennett, Lindley, Hollister, Gordon, & Springer, 1960) schizophrenics that tranquilizing drugs of the phenothiazine derivative group are associated with marked behavioral change, the exact manner in which the drugs produce these changes is not clearly understood. One hypothesis which has received general acceptance is that the behavioral change is secondary to a reduction of autonomic responses to anxiety-producing stimuli. Despite the general acceptance of this hypothesis, surprisingly few attempts have been made to test it experimentally on schizophrenic patients.

In contrast to the plethora of studies on behavioral effects of tranquilizing drugs, only a small number of studies have been reported on the effects of drugs on the autonomic nervous system response in schizophrenia. Particularly lacking has been research relating autonomic responses to emotionally arousing stimuli to drug status. The research which has been done (e.g., Mitchell & Zax, 1959) has suggested a reduction in autonomic responsiveness as a function of drug status; however, this research has been difficult in its interpretation for the following two reasons. First, a single autonomic response indicator (typically skin resistance) was used. This made it difficult to determine whether the reduction in responsiveness noted was entirely local (drying of the skin of the palm) or part of a general attenuation in autonomic nervous system activity. Second, the stimuli used were not of a class ordinarily believed highly potent in arousing anxiety in the schizophrenic (a loud tone). Clinical evidence (Arieti, 1955) suggests that the kind of stimuli which elicit anxiety in the schizophrenic are personal-social in nature, involving close interpersonal relationships, sexual cues, aggression, and self-assertive behavior. In the present study, the first of these difficulties was countered by the use of multiple measures of autonomic activity obtained concurrently to determine the generality of the drug effect. With respect to the second difficulty, it was felt that the effects of tranquilizing drugs could be best estimated using stimuli which were likely to maximize involvement and which dealt with thematic materials likely to mobilize anxiety in the schizophrenic. Therefore, tension-arousing films were selected as the stimuli, with psychophysiological responses to the films as the dependent variables. The basic hypothesis was that all indexes of autonomic reactivity recorded would show attenuation under the drug condition when compared to the individual's reactivity to a similar stimulus under the placebo condition.

---

[1] This study was supported by Grants MH 04720 and 08744 from the National Institute of Mental Health and by the Veterans Administration Center in Los Angeles through staff and facilities support. The authors would like to express their appreciation to Carol April, Paul Lewinsohn, Linda Miller, Granville Murphy, Marianne Osborn, and Charles Swearingen who assisted in data collection and analysis. Special thanks are also due the Health Sciences Computing Facilities at UCLA for computer time.

METHOD

Subjects. There were 40 chronic schizophrenics selected at random from the wards of the Brentwood Veterans Administration Hospital who served as subjects (*Ss*). The *Ss* were drawn from one ward at a time, assigned alternately to a drug or placebo group and, when all available patients had been selected from that ward, *Ss* were selected consecutively from another available ward. This procedure was continued until 21 drug *Ss*[2] and 19 placebo *Ss* were run in the experiment.

Basic Design of the Experiment. The experiment involved basically an *off-on* design. Each patient was withdrawn from his current medication and placed on a placebo for a period of time called the *withdrawal period.* During the withdrawal period, urine samples were collected upon arising three times a week. Samples were checked by a test (Forrest, Forrest, & Mason, 1961) for the presence of phenothiazine in the urine.[3] While this test is much less sensitive to the presence of thioridazine (Mellaril), it still provided a crude index of the amount of drug in the urine. A score of 0, or +1 following a 0, on this colorimetric test was identified as a "negative" reading, while a score of +2 or more (+6 maximum possible) was identified as a "positive" reading for purposes of this study.

Before beginning the experimental procedures, an *S* was required to have been on placebo for a minimum of 2 weeks and also to have shown two successive phenothiazine-negative urine tests (to assure excretion of prestudy medication). The mean time between drug withdrawal and the first experimental session was 18.3 days for the total patient sample. It should be recognized that the absence of phenothiazine in the urine does not indicate a complete absence of the drug in the body (Forrest et al., 1961). Clinical studies (Caffey, Forrest, Frank, & Klett, 1963) indicate that regressions in clinical behavior continue to occur for some time after the laboratory procedures fail to detect any substantial amount of phenothiazine. This strongly suggests that a certain percentage of the drug is retained long after active drug withdrawal. Thus, *Ss* at the time of first testing could be more accurately characterized as being at a *minimal,* rather than at a *nondrug* level. The 2-week minimum period between withdrawal and testing represented essentially a compromise between the ideal experiment and certain biases which would have been introduced by a longer withdrawal period. Previous research in our laboratory has indicated that when an experiment extends for more than a 30-day period, a substantial loss in *Ss* occurs through discharge or regression to a more acute state which requires readministration of an active drug. Further, Chapman (1963) has reported that patients who show this behavioral disorganization earliest represent a selective group in terms of manifesting a more severe thought disorder, whose absence from the sample introduces limitations in the generality of drug-study findings. Therefore, the advantage of a minimal loss in the *S* sample (actually no patients were lost) was chosen over the disadvantage of testing patients with less than complete drug withdrawal.

Drugs. The experimental medication in this study was thioridazine (Mellaril: Sandoz).[4] This medication was chosen as being fairly typical of the more commonly used phenothiazine derivatives. It resembles chlorpromazine (Thorazine: Smith, Kline & French) in the clinical indications and in its clinical effects. However, it does not customarily require additional medication for the control of extrapyramidal symptoms (pseudoparkinsonism), which would introduce added complications in a psychophysiological drug study.

The drug was administered in 100-milligram tablets. The placebo medication was provided by the manufacturer and was identical in appearance and taste to the active medication. The initial dosage was one tablet four times per day. The ward physician was allowed to raise or lower this dosage within moderate limits according to apparent clinical requirements, but this rarely occurred. The drug was identified by a numerical code. The ward physician, nurses, and research assistants on the project who actually saw the patients did not know which was active and which was placebo medication.

First Experimental Session. When *S* had satisfied the urine test criterion and at least 2 weeks had passed from the initiation of active withdrawal, he was asked to report to the psychology laboratory (which was located adjacent to the hospital EEG laboratory). The *S* was told that the experimenters (*Es*) were going to take some measures "to see how his treatment was affecting him." Following this, the patient was taken into a sound-attenuated 9 × 9 foot air-conditioned chamber,[5] shown the psychophysiological leads, and given a brief introduction to what was measured by each lead. The *S* was seated in an easy chair, pickups were attached and *S* was told to relax for a 10-minute period, after which a film would appear on the screen, which he was requested to watch. The *E* then left the room and the 10-minute adaptation period was begun which was designed to acclimatize the patient to the attachments and experimental room. Following this period, during which psychophysiological measures were recorded continuously, the projector was started and a 24-minute sound film was run. After the film, the patient was unhooked from the equipment and interviewed concerning his comprehension and reaction to the film. He was taken to an adjoining room where other psychological procedures, including a word-association test and, where possible, The Perceptual Defense Battery (Goldstein, Jones, & Kinder, 1964), were carried out.

---

[2] The authors' original plan was for 20 drug and 20 placebo patients but due to a clerical error the assignment ended up 21 to 19.

[3] The authors are indebted to Donald Parkhurst, Chief, Clinical Laboratory, Brentwood Veterans Administration Hospital and his staff for running the phenothiazine urine tests.

[4] The authors are indebted to the Sandoz Company for providing them with drugs and placebo.

[5] Manufactured by Industrial Acoustics Corporation, Santa Monica, California.

Second Test. Procedures for the second experimental session, when *S* viewed the alternate film, were identical to those of the first session. The *Ss* were run on the second testing day at the same time of day that they were run on first testing. Half of the *Ss* were run in the morning and half in the afternoon.

Movies Used. The films used in the experiment were both made by Alfred Hitchcock as television programs.[6] The films were presented in counterbalanced order with one half of the *Ss* receiving one film first and the other half, the other film first. The two films were titled *Man From the South* (*MS*) and *No Pain* (*NP*). The basic plot of *MS* involves a man who has bet his pinky finger that his cigarette lighter will fire successfully 10 times in a row. The film builds steadily in tension to a scene in which a somewhat sadistic gambler stands with a meat cleaver poised over the bound finger of the hero, who is going through the sequence of lighting his lighter. The film, which was judged quite arousing by all who saw it, dealt with stimuli likely to be threatening to the schizophrenic such as aggression, mutilation, and possibly castration anxiety. The plot of *NP* dealt with a completely paralyzed man who was required to spend all of his time in an iron lung, totally dependent upon his vivacious young wife. As the story develops, there are a number of arousing scenes, one a flashback of a sexual nature in which the man is shown as a vigorous and virile person making love to his wife; the film then returns to the man paralyzed and impotent. Later in the film, it becomes obvious that the wife is trying to kill her husband and there is uncertainty as to whether or when the wife will permanently disconnect the iron lung. Again, this film presented a number of themes potentially disturbing to the schizophrenic such as infidelity of the wife, sexual arousal, and the contrast between early masculine virility and later impotence and helplessness. Fortunately, analyses of psychophysiological responses of the total patient groups indicated no difference between movies in average level of reactivity so that they could be treated as relatively equivalent tests in their arousal potential.

Assignment to Drug or Placebo Status. Prior to the first movie viewing, *Ss* were assigned alternately to the Drug or Placebo groups. Following the first movie test session, the placebo group continued to receive the same placebo medication administered during the withdrawal period. Within 12 hours of the first test session, the drug group began to receive the active drug medication.

During the period following the assignment to drug or placebo status, the phenothiazine urine test was done three times a week on alternate days as a check on whether patients were truly ingesting the pills. Second testing occurred when a minimum of 7 days had elapsed. In addition, two consecutive positive urine tests were required for retesting in the drug groups and two consecutive negative tests were required for retesting the placebo group. As indicated previously, the range of scores obtained on this test with this particular phenothiazine was very narrow; however, the readings did correlate with drug or placebo status such that there was virtually no overlap between the groups on the second testing day. The mean number of days between first and second testing was 8.8 days.

Psychophysiological Measures Recorded. Throughout the 10-minute rest period and the 24-minute film period, the following measures were continuously recorded by an Offner Type R Dynograph recorder: (*a*) *Skin resistance* with palm-to-palm attachment, using cup-type (2 centimeters in diameter) zinc electrodes, zinc sulfate paste, and a 40-microampere Darrow-type bridge[7] with DC coupling to the polygraph recorder. (*b*) *Heart rate,* with electrodes attached to each side of the chest using a cardiotachometer (see Footnote 7) with AC coupling to the polygraph recorder. (*c*) *Finger-Pulse Volume* using a finger cup on the left index finger with a closed pneumatic connection to a plethysmograph transducer (see Footnote 7) with AC coupling to the polygraph recorder.

RESULTS

The basic design of this study required some comparison of indexes of psychophysiological reactivity for the two film viewings. This required that for each psychophysiological variable some summary measures be devised which could reflect the important parameters of a 24-minute reaction period; since these summary measures varied for each variable, the results will be discussed separately for each psychophysiological variable.

---

[6] Special thanks are due Alfred Hitchcock Productions and Revue Studios, Universal City, California, for generously providing the authors with films on an extended loan basis.

[7] Manufactured by Electro Medical Engineering, Burbank, California.

Skin Resistance. Admittedly, there are an infinite number of indexes which can be derived from a continuously recorded skin-resistance curve. There was particular interest in two dimensions of the skin-resistance response, the long-term shifts in skin-resistance level upward or downward from a basal level and the relatively discrete changes (GSRs) occurring at critical scenes in the movie. The independence of these dimensions is strongly supported by Spiesman, Osborn, and Lazarus (1961) in their parametric study of skin-resistance measures. Two measures of long-range changes were of particular interest: (*a*) The net activation-relaxation score of the individual, and (*b*) the total range of response from maximum to minimum tension level. With regard to the former, it is obvious that during response to a complex stimulus such as a film there are two general processes operating, one involving the gradual adaptation of *S* to the film viewing experience which should be accompanied by a rising skin resistance and a second process of reactivity to specific tension-arousing scenes which should be accompanied by a declining skin resistance. Therefore there was interest in whether the drops in skin resistance (activation) were greater than or equal to the rises in skin resistance above resting level (relaxation) due to long-range adaptation. The mean level of response was not chosen as an index of arousal despite its high loading on Cluster II in the Spiesman et al. (1961) research as it is too insensitive a measure for a study of this sort. It provides a crude index of stress response when two stimuli widely different in tension potential are compared, but is not sufficiently sensitive to reveal individual differences in reactivity, a conclusion also reached by the Lazarus group in later papers (Spiesman, Lazarus, Mordkoff, & Davison, 1964).

The skin-resistance curves for the 24-minute film period were read for the following parameters and converted to log conductance units: (*a*) *Beginning conductance:* this was the highest conductance appearing during the 30 seconds following the start of the projector and appearance of the film on the screen (orienting response to projector or onset of film). This measure, rather than one of the resting scores, was used as a basal index for correcting the later measures as it was noted that *S*s frequently failed to recover fully from their orienting response such that this became the basal level of an alerted organism against which all later changes to emotional stimuli could be more appropriately compared. (*b*) *Highest conductance* during the film. This reflected the point of maximum activation during the film following the initial 30-second interval. (*c*) *Lowest conductance* during the film. This reflected the point of maximum relaxation during the film. (*d*) *The average of the three largest GSRs* within a 10-second period occurring at any three points during the course of the film. Certain summary indexes were derived from *a, b,* and *c* measures as follows:

1. *Net Activation-Relaxation:* A measure which reflected whether *S* showed a greater rise or drop in conductance from his basal level during the course of the film. This score was computed as follows: (highest conductance − beginning level) minus (beginning level − lowest conductance). This *net* score was computed such that a plus score reflected greater activation than relaxation while a minus score had the opposite meaning.
2. *Total range of skin-conductance change.* This measure reflected the total range of skin-conductance change regardless of the nature of the change. This was computed as follows: (highest conductance − beginning) plus (beginning − lowest conductance).
3. *Average of three largest GSRs* which was described above.

The means for the three indexes for first and second testing are presented in Table 1. It can be seen in Table 1 that the changes from first to second viewings for drug and placebo groups vary with the index used. The measure of net activation-relaxation (highest − beginning) − (beginning − lowest) fails to indicate any significant effect due to drug status; however, the total range and the average of the three largest GSRs indicate a diminished responsiveness for the drug group and unexpectedly an *enhanced* responsiveness for the placebo group on the second testing.

The skin-conductance data were analyzed by analysis of covariance in which the initial

TABLE 1

MEAN SKIN-CONDUCTANCE CHANGES (LOG CONDUCTANCE CHANGE) ON 3 MEASURES FOR DRUG AND PLACEBO GROUPS

Measure	First showing	Second showing	*F*
Net activation-relaxation (highest − beginning) − (beginning − lowest)			
Drug	.032	.041	2.29
Placebo	.008	−.031	
Total range (highest − beginning) + (beginning − lowest)			
Drug	.205	.102	12.16*
Placebo	.175	.185	
Average 3 GSRs			
Drug	.061	.022	15.48*
Placebo	.047	.062	

* $p < .01$ for 1 and 38 *df*.

testing score was used as a covariate to adjust the second testing score. However, while all *F*s reported in Table 1 are from covariance analyses, the means represented there are not adjusted means but are the actual means for each session. The drug group and the placebo group do not differ significantly on total range and average of three largest GSRs on the first testing, indicating that the difference between the groups appears only after the active drug agent has been introduced. Thus, there is a dampening of response in the drug group on two out of three measures and slightly greater responsiveness in the placebo group.

Covariance analysis was also done in a manner which permitted analysis of order effects (*MS* or *NP* on first showing). Order of film presentation was not significant for each of our three indexes of skin-resistance response (*F*s = 0.14, 0.30, and 1.33, respectively for measures as listed in Table 1), and the interaction of order of film presentation and drug status (*F*s = 0.18, 0.04, 0.8 for the three indexes in Table 1) also was not significant. Thus, the reduction in responsiveness noted under the drug condition was independent of the content of the film viewed in the second session.

Heart-Rate Data. The data for heart rate were read and analyzed in a different fashion from skin resistance. Research using films (Malmstrom, 1964) indicates that heart rate does not follow the events of a film with nearly the precision of skin resistance. The various automatic homeostatic mechanisms which operate to bring deviations back into normal range as well as the cyclical relationship to respiration interact with heart-rate changes to a film to produce a given record of response. Therefore, the type of measures used for skin resistance was not considered appropriate for heart rate. Instead, following Malmstrom's suggestion, a method of moving averages was used to identify the period of maximum heart-rate acceleration. This method was used to average out short-run homeostatic and respiratory cycles. This type of measure is not directly comparable to the measures used for skin resistance but appears to be more sensitive to the cyclical nature of the heart-rate response, a quality not present in skin resistance.

Six 1-minute periods were chosen so as to span six potentially emotional scenes in each movie such that each period began 30 seconds prior to the emotional scene and continued into the high point of the scene. Fortunately, the two movies were so constructed that the same six 1-minute time intervals could be used in reading records for both films. This means that the period of time from onset of the projector was approximately the same for each 1-minute period read for *MS* and *NP*. Within each 1-minute segment, heart rate was read continuously. The data obtained were then treated as a continuous series, and moving averages were computed to smooth out irregularities in the time series. The moving averages were computed by taking consecutive averages of beats per minute for each set of 30 adjacent beats (the means of Beats 1–30, 2–31, 3–32, 4–33, etc.). The largest moving average obtained was treated as a stable index of the individual's peak response of heart-rate acceleration during the six 1-minute series and was used as a summary index of peak reactivity to the film.

Adjustment of these data for basal level is particularly critical in a study of this type as the phenothiazine drugs have been noted to cause tachycardia in a substantial number of patients. Thus, any estimates of variation in reactivity would have to be made against variations in basal level from Session 1 to Session 2. The choice of basal or initial level is not a simple matter. While the resting-level data prior to the onset of the film would be a conventional analogue of the typical prestimulus measure, it was obvious from post-session interviews that many schizophrenics were extremely fearful during the resting period and apparently apprehensive that something painful was about to happen to them. Many doubted that a film would be shown and were distrustful on this score. Frequently, following the onset of the projector, a brief heart-rate acceleration was followed by a continuous decline in the average heart-rate level. Bursts of acceleration could be seen from this declining level, but frequently the overall trend was descending to a level lower than the "resting" level. Thus, it was uncertain whether the prestimulus level or the lowest level achieved during the film was the more appropriate basal level for adjusting the reactivity scores. Since no rational basis existed for choosing between them, the heart-rate data were analyzed twice, once for each basal level. The correction for basal level used was that recommended by Lacey (1959) in which *S*'s corrected score is his deviation from the regression line of basal level on reactivity. The scores generated are deviation scores which are plus or minus in sign depending on whether *S* is more or less reactive than would be predicted on the basis of his basal level. A corrected score was computed for each *S* for each experimental session and the difference between the two scores was obtained such that a plus score indicated greater reactivity on the second showing and a minus score, less reactivity. Of the original sample of 40 *S*s only 34 provided two sets of heart-rate data which were deemed satisfactory for analysis. The mean difference scores in which the premovie resting level was used as the basal level to predict reactivity were -3.05 for the drug group and $+0.93$ for the placebo group. Analysis of variance of these difference scores revealed a between-groups F of 5.01 ($df = 1/32$) which is significant at the .05 level. The mean difference scores where the lowest level during the film was used as the basal level to adjust reactivity were -5.13 for the drug group and $+1.13$ for the placebo group. The F from the analysis of variance for these difference scores was 6.99 ($df = 1/32$), which is also significant at the .05 level. Thus, the drug group shows less reactivity in heart rate on the second showing while the placebo group, as in the findings reported above for skin-resistance data, shows slightly greater reactivity on the second showing, independent of the basal measure used. Once again we failed to find a significant effect for type of film or the interaction between film order and drug status.

The data were also analyzed using a measure of variability in heart rate, a dimension found independent of mean level and fastest period by the Spiesman et al. (1961) cluster analysis. Using the previously cited six 1-minute continuously read data, standard deviations based on 430 beats were computed for each individual. Since a scatter plot failed to reveal a correlation with initial heart-rate level, the data were not corrected in any fashion. Once again the difference between first and second viewing was computed such that a minus equaled lowered reactivity on the second showing. The mean difference score for the drug group in standard deviation of heart rate was -1.50 and for the placebo group -0.07 ($F = 4.18$, $p < .05$, $df = 1/32$) which is statistically significant. Thus, the variability in heart rate in the drug group is less during the second film viewing and shows little or no changes in the placebo group, supporting the analysis of fastest moving average.

Finger-Pulse Volume Data. The Plethysmograph data were analyzed in a fashion somewhat analogous to the skin-resistance data. Three measures were used: (*a*) *Beginning – Maximum constriction level* during the film. For this measure, the beginning level was used as a predictor to adjust this difference score in obtaining a corrected Lacey score. (*b*) (*Beginning – Maximum constriction level*) + (*Maximum dilation level – Beginning level*). This measure reflects the maximum range in finger-pulse volume during the film. Once again beginning level was used as a predictor to obtain corrected levels of the complex difference score. (*c*) *The average of the three largest con-*

strictions during the film. A constriction was defined as a marked change in pulse amplitude within a 10-second period. The preconstriction level just prior to each individually selected constriction was used to obtain a corrected score for that constriction and the average of the three greatest corrected scores was then obtained for each *S*. All corrections of finger-pulse data used the Lacey (1959) procedure. Table 2 presents the mean difference score on each measure (second minus first film viewing). It can be seen in Table 2 that, while two of the three measures follow the pattern noted earlier for skin resistance and heart rate of lesser reactivity in the drug group and greater reactivity in the placebo group, none of the means differ reliably from each other. However, a more detailed analysis of the data is instructive here. Only 6 of the 17 drug patients showed greater reactivity on the second showing, while 12 of the 18 placebo *S*s with usable plethysmograph records showed greater reactivity on the second showing. These frequencies would tend to support the general hypothesis. Furthermore, in each group the scores which go against the group trend are very large in magnitude and it is this phenomenon which increases the error term markedly and thereby reduces the level of significance of the results.

Exploratory Study of Drug Effects on Associative Processes. A supplementary aspect of the present study was an investigation of changes in the nature of word-association test responses in drug and placebo groups. It was hoped that a study of this class of response, so closely allied to thinking disorder in schizophrenic patients, would provide an index of changes in thought patterns which might accompany reduction in autonomic responsiveness. No causal hypothesis was entertained on the relationship between these two response systems as it was considered entirely possible that reduction in autonomic responsiveness could result from, cause, or be entirely independent of changes in thought patterns. The exploratory study was conducted on the final 20 patients run in the study. Each patient was administered a 50-item word-association test immediately following the first movie viewing and was readministered the same test following the second movie viewing. The test was scored according to a scoring system described elsewhere (Goldstein & Jones, 1964) and involved two clusters of response categories located in that study, one cluster based on the temporal aspects of responses and a second category derived from measures of thinking and memory disturbance. The temporal scores were: (*a*) the number of fast reaction times; *No. FRT* (2 seconds or less); (*b*) the mean reaction time of all 50 responses, $\bar{X}RT$; (*c*) the variance in reaction time of all 50 responses, $\sigma^2 RT$. The thinking and memory disturbance scores were: (*a*) the number of Distant responses (*D*). This category is described more fully in the earlier publication and refers to responses in which the associative connection between stimulus and response are remote but possess some logical connection; (*b*) Austistic responses (*A*), essentially a more severe form of the *D* response but one in which no logical connection between stimulus and response can be discerned by an external observer. An example of this category would be the stimulus-response sequence, *fight-tree;* (*c*) Number of memory disturbances (*M*). This score is based on the number of responses reproduced incorrectly on the second administration of the word-association test. (After completing the list the test was immediately administered a second time with instructions to *S* to give the same response as on the first administration.) The latter category of responses was found in a previous investigation to cor-

TABLE 2

MEAN DIFFERENCE IN CORRECTED REACTIVITY SCORES FOR 3 MEASURES OF FINGER-PULSE VOLUME CHANGES

Measure	Drug ($N=17$)	Placebo ($N=18$)	F
(Beginning − Maximum constriction)	−0.10	−0.14	0.01
(Beginning − Maximum constriction) + (Maximum dilation − Beginning)	−0.27	0.28	1.45
Average of 3 constrictions	−0.21	0.22	1.60

TABLE 3

ADJUSTED MEANS AND *SD*s ON WORD-ASSOCIATION VARIABLES FOR DRUG AND PLACEBO GROUPS

Measure	Group	*X*	*SD*	*F*
Tempo measures				
No. FRT	Drug 10	17.95	2.53	1.00
	Placebo 9[a]	14.27	2.67	
$\bar{X}RT$	Drug 10	3.33	0.43	0.683
	Placebo 9	3.85	0.45	
$\sigma^2 RT$	Drug 10	6.99	2.79	0.163
	Placebo 9	8.64	2.94	
Thinking and memory measures				
D^b	Drug 10	46.20	3.48	2.69
	Placebo 9	54.55	3.67	
A	Drug 10	2.53	0.67	5.61^a
	Placebo 9	4.84	0.70	
M	Drug 10	14.89	1.85	1.42
	Placebo 9	11.68	1.95	

[a] One placebo patient refused to take the test on the second administration reducing the placebo *N* to 9.

[b] The *D* score is based on a weighting formula in which two classes of responses are considered, mildly distant and moderately distant, with the latter assigned twice the weight of the former. Thus, the *D* score can be larger than the number of stimulus words.

relate moderately in a negative direction with behavior ratings of social contact by aides and nurses on the ward, with high scores associated with poor social contact. The data were treated by analysis of covariance in which the first session score on a variable was used as a covariate for the second session scores. The adjusted means and standard deviations for each variable are presented in Table 3. It can be seen in Table 3 that there is only one variable on which a significant shift in associative responses can be noted and that is *A*, the most severe form of idiosyncratic response. The drug group's adjusted mean is significantly lower than the placebo mean; in addition it is lower than its initial mean. The placebo group shows an increase in autistic associations above its initial mean. These results must be interpreted with caution as they are only derived from half of the total sample; yet, it is interesting to note a parallel to the psychophysiological results in that the drug group shows a reduction in this response class while the placebo group shows an increase in this same class.

DISCUSSION

The present results support the hypothesis that ingestion of a tranquilizing drug of the phenothiazine group is associated with a reduction in autonomic reactivity to an anxiety-linked stimulus in schizophrenic patients. The fact that this trend is apparent at a significant level on two indexes of autonomic reactivity and shows a similar trend on the third suggests that the present findings are not the result of peripheral changes (e.g., skin dehydration) but are more likely due to a general dampening of autonomic nervous system responsiveness. While large variations in the amount of reduction in responsiveness can be noted among the patients, some reduction in skin resistance and heart-rate reactivity was noted for the great majority of drug patients, while an increase in reactivity or no change was noted for the large majority of placebo patients. One link in the chain of evidence which remains to be explored is the relationship between the magnitude of attenuation in autonomic reactivity in a drug patient and the degree of alteration in clinical behavior during a comparable period. It seems probable, but largely unproven, that correlations between these two systems of behavior should exist such that large reductions in autonomic reactivity accompany and possibly precede reductions in agitation on a clinical level.

The reduction in autonomic reactivity in the drug group raises some question of the mediating mechanism for this attenuation. Did some change in cognitive appraisal of threat take place as a result of the drug which inhibited emotional arousal or, have cognitive appraisals of events remained constant but the associated emotional response dampened in some fashion? As the experiment progressed, the theoretical importance of these points became clear, and some attempt was made in the final run of 21 patients (12 drug and 9 placebo) to obtain some index of *S*'s comprehension of the specific events in the film. As mentioned earlier, each *S* was interviewed following each film viewing regarding his reaction to the film, but only on the final series of 21 *S*s was a systematic rating of response done. A 5-point scale of comprehension[8] was constructed which scaled the events in the film in order of importance to plot comprehension. The events mentioned were listed under each scale point and *S*'s rating was the number of events clearly perceived according to his interview. The ratings were made by one of the research assistants who ran *S*s through the movie viewing and who was also "blind" regarding *S*s' drug status. On these ratings, the higher the score, the greater the comprehension of the plot. The mean ratings of comprehension for the drug *S*s were: first time = 3.42, second time = 3.42. For the placebo *S*s: first time = 3.33, second time = 3.11. None of the means differed significantly from each other. Thus, there is actually no difference in the two groups in their initial level of comprehension nor in their comprehension of the second film. These data, while drawn from only half of our sample, do suggest that the autonomic changes noted are not paralleled by systematic changes in the clarity of perception of events in the film. Thus, if the drug or placebo status is associated with some systematic shifts in cognitive appraisals, the shifts involve more subtle factors than comprehension of the plot as determined by the rough ratings used here. Accordingly, our limited data appear congruent with the analysis of Himwich (1960) of the effects of chlorpromazine administration on response to painful stimuli in the rabbit. He states

---

[8] The authors are indebted to Paul Lewinsohn for developing the scale.

> when chlorpromazine is given, the reticular formation is blocked and no longer shows the alerting response. This inhibition of the reticular formation does not prevent impulses evoked by a painful stimulus from traversing the lemnisci and awareness is thus achieved by way of the cortical mechanisms. But the severity of the associated distressing emotional accompaniment of the stimulus is greatly mitigated [p. 65].

The findings in the pilot study on association responses, if replicated on another sample, raise some interesting theoretical questions regarding the relationship between remote, idiosyncratic associations, and level of autonomic responsiveness. One possibility, suggested by the work of Mednick (1958), is that remote associative responses are anxiety-reducing responses which gradually become highly practiced because they reduce tension. If a phenothiazine drug reduces anxiety it may in essence do for the individual what the remote associations did formerly; thus, under the drug condition less of these responses are needed to avoid anxiety arousal. This would also account for the increase in remote responses in the placebo group where an increasing anxiety level might stimulate increased use of low probability associates. Another possibility is that the drug affects associative processes primarily and that the reduction in autonomic reactivity is due to the suppression of primary-process material in awareness. A third possibility is that the phenothiazines affect associative processes and autonomic responses concurrently and that these parallel response decrements mutually reinforce each other over a period of time to produce more appropriate behavior. Certainly, more research designed to test these various alternatives will do much to identify the locus of phenothiazine drug action on a psychological level.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955. Ch. 3.

CAFFEY, E. M., FORREST, I. S., FRANK, T. V., & KLETT, C. J. Phenothiazine excretion in chronic schizophrenics. *American Journal of Psychiatry,* 1963, **120**, 578–582.

CASEY, J. F., BENNETT, I. F., LINDLEY, C. J., HOLLISTER, L. E., GORDON, M. J., & SPRINGER, N. N. Drug therapy in schizophrenia: A controlled study of the relative effectiveness of chlorpromazine, promazine, phenobarbital and placebo. *Archives of General Psychiatry,* 1960, **2**, 210–220.

CHAPMAN, L. F. The problem of selecting drug-free schizophrenics for research. *Journal of Consulting Psychology,* 1963, **27**, 540–542.

FORREST, F. M., FORREST, I. S., & MASON, A. S. Review of rapid urine tests for phenothiazine and related drugs. *American Journal of Psychiatry,* 1961, **118**, 300–307.

GOLDSTEIN, M. J., & JONES, R. B. The relationships among word association test, objective personality test scores and ratings of clinical behavior in psychiatric patients. *Journal of Projective Techniques and Personality Assessment,* 1964, **28**, 271–279.

GOLDSTEIN, M. J., JONES, R. B., & KINDER, M. I. A method for the experimental analysis of psychological defenses through perception. *Journal of Psychiatric Research,* 1964, **2**, 135–146.

HIMWICH, H. E. Biochemical and neurophysiological action of psychoactive drugs. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960.

LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.

MALMSTROM, E. J. An improved method of heart rate measurement: Correlations with skin conductance and psychological reaction under stress. Paper read at Western Psychological Association, Portland, Oregon, April 1964.

MEDNICK, S. A learning theory approach to research in schizophrenia. *Psychological Bulletin,* 1958, **55**, 316–327.

MITCHELL, L. E., & ZAX, M. The effects of chlorpromazine on GSR conditioning. *Journal of Abnormal and Social Psychology,* 1959, **59**, 246–249.

PSYCHOPHARMACOLOGY STUDY CENTER GROUP. Phenothiazine treatment in acute schizophrenia. *Archives of General Psychiatry,* 1964, **10**, 246–261.

SPIESMAN, J. C., LAZARUS, R. S., MORDKOFF, A., & DAVISON, L. Experimental reduction of stress based on ego defense theory. *Journal of Abnormal and Social Psychology,* 1964, **68**, 367–380.

SPIESMAN, J. C., OSBORN, J., & LAZARUS, R. S. Cluster analysis of skin resistance and heart rate at rest and under stress. *Psychosomatic Medicine,* 1961, **23**, 323–343.

(Received April 4, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 345–353

FAMILY INTERACTION WITH SCHIZOPHRENICS AND THEIR SIBLINGS

SHLOMO SHARAN (SINGER)[1]

Albert Einstein College of Medicine, New York

24 families, each including both parents, a schizophrenic, and a nonschizophrenic adolescent, were subdivided into a parent-patient and a parent-nonpatient triad and asked to solve collectively the questions from the Comprehension and Similarities subtests of the Wechsler-Bellevue Intelligence Scale. The groups' recorded discussions were compared for problem-solving efficiency, mutual support patterns, and parent-child sex-role alignments. Contrary to predictions, the 2 triads displayed equal efficiency; parents supported both children equally; fathers and mothers were equally dominant. Patients were more supportive of their parents than were siblings, while parental discord was more prominent in patient than in nonpatient groups.

A perspective on schizophrenia is emerging in which the illness is viewed as symptomatic of pathological interaction patterns in the entire family (Bowen, 1960; Lidz, 1963). The empirical foundations for these family theories rest upon data from sources such as family therapy sessions, interviews, and the observation of families living on the wards of hospitals. However, there are to date only a handful of studies undertaking an experimental exploration of the complex issues surrounding the family life of schizophrenic patients.

The present experiment selected three aspects of current family theories for investigation. These were: (*a*) impaired problem solving within the family; (*b*) the formation of rival intrafamilial dyads; and (*c*) parental role-reversals and the crossing of sex-generation boundaries. If such behavior patterns are indeed manifest, are they associated exclusively with a schizophrenic child, or do they also pervade family interaction in the presence of a nonschizophrenic sibling? Are the normal children also involved in the same pathological family interaction as their schizophrenic siblings? If the family is schizophrenogenic, how does it happen that all of the siblings are not diagnosable as schizophrenic patients (Bateson, 1960; Lidz, Fleck, Alanen, & Cornelison, 1963)? Clearly the presence of a normal child in the same family with a patient presents a challenge to the family theories. It also provides a unique opportunity for the systematic differentiation of those patterns of family behavior which may be specific to the parent-patient interaction in contrast to the parents' interaction with their nonpatient-child. The major task of the present study was directed toward precisely this comparison of family interaction with a patient- and a nonpatient-child. As such, it was a study of intrafamilial relationships and did not purport to provide information about etiological factors in schizophrenia.

Several observers maintain that communication in schizophrenic families serves defensive needs and possesses little instrumental utility for dealing with the environment (Fleck, Lidz, & Cornelison, 1963; Haley, 1959). On the basis of these observations, it was hypothesized that the family triad involving the patient should manifest impaired problem solving when compared to the triad with the nonpatient-child. Hence the hypothesis:

If disturbed communication in the parent-patient triad impairs problem solving, then the parent-patient triad should solve problems less efficiently than the parent-sibling triad.

The second parameter of pathological interaction which was selected for study in this experiment concerns the competition between the parents for the allegiance of their patient-child. It has been noted that when the family is split into rival dyads, the interaction of its members is marked by chronic conflict, and the parents do not form a mutually constructive relationship with the patient for purposes of problem solving (Bowen, 1960; Lidz, Cornelison, Fleck, & Terry, 1956). These observations led to the expectation that, not only would there be considerable expression of mutually deflating and critical remarks in the family during the course of its problem-solving efforts, but that such destructive interaction should be more apparent in the patient than in the nonpatient triad. This expectation was supported by case reports of families with both a patient and a nonpatient-child who resided on a hospital ward for extended periods. The impression was that the intrafamilial conflict was confined largely to the parent-patient interaction (Brodey, 1959).

---

[1] Now at Tel-Aviv University, Israel.

Two interview studies have addressed themselves to the normal child in the schizophrenic's family (Lu, 1961, 1962; Prout & White, 1956). These authors concur that the patients were not singled out for systematic scapegoating or abuse by parents, nor were they subject to unusual physical or psychological traumata. Rather, it appears that the patients were always more passive, more sensitive, and less sociable than their normal siblings. By contrast the siblings always were less submissive and more independent than the patients. These studies indicated that the observations of some family theorists regarding the parental denigration of patients still require systematic, experimental investigation. Therefore, it was hypothesized that:

If the parents do not form a supportive relationship with the patient for purposes of problem solving, then the parents should be less supportive of the patient than they are of his sibling.

Finally, family studies have noted a reversal of roles between the parents of schizophrenic patients, as well as a breakdown in the integrity of the sex-generation boundaries between parents and patients. Fathers of male patients are allegedly passive and aloof from family affairs, while by contrast the mothers are reported to be the more aggressive parent, maintaining an engulfing, symbiotic relationship with their patient-sons. With female patients, the parental roles rotate, with the mothers aloof and withdrawn and in competition with their daughters. On the other hand, the fathers are aggressive, even tyrannical toward their wives, and they maintain an intensive, seductive relationship with their patient-daughters (Lidz, 1963).

Several experimental studies have examined the question of parental role-reversal. Farina and Dunham (1963) reported that the fathers of the patients with "good" premorbid adjustment, as well as the fathers of the control subjects (*S*s), tended to dominate in their families, while the only tendency for maternal domination appeared in the families of patients with "poor" premorbid adjustment, although this tendency did not achieve statistical significance. Caputo's (1963) results also did not support the role-reversal hypothesis: "bilateral rather than unilateral expression of hostility between those parents seemed to be characteristic . . . both parents . . . contribute to the so-called schizophrenic character of the family [p. 354]." Cheek (1964) concluded that her results were inconclusive on the issue of parental role-reversal. Mothers appeared to be close to their sons only in contrast to the more peripheral position of the father, but these mothers were also cold, withdrawn, and expressed very little support of their patient-sons. Haley (1962) found that the parents of schizophrenic patients tended to interact more with each other than with their child. This finding might be consistent with the hypothesis that the parents form a closer alliance with the normal than with the patient-child. Lennard, Beaulieu, and Embrey (1965) interpreted their findings as consistent with the role-reversal hypothesis. Family experiments have thus yielded conflicting results regarding role-reversal in the families of schizophrenic patients.

For the purpose of experimental clarity, the role-reversal hypothesis was subdivided into two parts: (*a*) the interparental relationship; and (*b*) the parent-patient relationship (Lidz, 1963). In respect to the interparental dyad, the observations reported by family theorists led to the expectation that the more active and dominant parent should be of the opposite sex of the patient. In a problem-solving situation, the dominant parent's answers should appear more often as the group's solution to the task than the answers of the parent who is of the same sex as the patient. Also, the dominant parent should be expected to participate in the family discussion more than the less assertive parent. Concerning the confusion of sex-generation boundaries, the dominant parent was expected to maintain a more intense and supportive relationship with the patient than the less dominant, same-sex parent did. The hypotheses, therefore, were as follows:

1. If there is a breakdown of parental sex-linked roles, then the patient and dominant parent should be of the opposite sex; and,

2. If there is a breakdown in the parent-child sex-generation boundaries, then the patient and opposite-sex parent should have a more mutually supportive relationship than the patient and same-sex parent.

METHOD

Subjects

The families of 24 schizophrenic patients, each family including both parents and a normal child in addition to the patient, were selected for study. All families were tested in the hospital while the patient member was hospitalized with a diagnosis of schizophrenia. The hospital staff's diagnosis was accepted as the criterion for including the patient in the category of schizophrenia. None of the patients or siblings had histories of organic impairment or mental retardation, nor had any of the patients been hospitalized for mental illness prior to adolescence, and all patients had been hospitalized for 1 year or less. All normal siblings were functioning adequately in school or on jobs and had no history of psychiatric treatment.

The families were divided into four groups according to the sex of patient and sibling. Each group consisted of six families with the following sex alignments of the children: male patients, female siblings; male patients, male siblings; female patients, male siblings; female patients, female siblings. The mean age of each category of *Ss* was as follows: patients, 19.8 years (*SD* 2.4); siblings, 17.2 years (*SD* 2.6); fathers, 50.2 years (*SD* 4.1); mothers, 47.2 years (*SD* 5.0). All patients and siblings were unmarried and residing at home with both parents, except for the period of the patients' hospitalization. Thus, both in age and in marital status, these groups were more homogeneous than in previous studies. In addition, it was more likely that these people still constituted a family, in terms of fulfilling typical parent-child roles, than in research involving elderly parents with married children who had spent years away from home. All families were Caucasian, 15 of Jewish, and 9 of Catholic religious background.

Procedure

The Interaction Testing Technique devised by Roman and Bauman (1960) was adopted as the research instrument. As used here, the technique consisted of the Comprehension and Similarities subtests of the Wechsler-Bellevue Intelligence Scale, Forms I and II, providing two comparable sets of standardized questions for use with the parent-patient and parent-sibling triad.

The testing procedure was as follows: All four family members were seated in such a way that they could not see each other's paper. Each one was given a mimeographed copy of the 44 questions from both forms of the two Wechsler subtests. The questions were combined so as to retain their progressive order of difficulty, that is, Question 1 from Form I was followed by Question 1 from Form II, etc. When everyone had answered all the questions in writing, the patient or sibling was asked to leave the room, the sequence of which child left first being alternated. The remaining triad was then given one copy of the questions from only one form of the Wechsler subtests, and they were asked to discuss the questions among themselves and formulate a response acceptable to the entire group. The examiner did not participate in and did not respond to questions during the group discussions, all of which were tape recorded. The instructions were as follows:

> I am going to readminister some of the questions which you just answered individually. The purpose is to determine whether the group can do better together than each person was able to do alone. Please discuss each question among yourselves and agree upon the best answer. Try to have everyone participate in the discussion, and have one person write the group's final answer on this paper. There is no time limit.

When the questions from subtests of one Wechsler form had been answered, that child left the room and the same procedure, using the alternate Wechsler form, was followed with the other child. Testing sessions with a family ranged from 1 to 2½ hours, with the mean session being approximately 1½ hours.

Measures

The written responses were scored according to the standard criteria for the Wechsler-Bellevue Intelligence Scale (Wechsler, 1944, 1946). From these scores a maximum and an actual score were calculated. The maximum score was obtained by selecting the highest score (0, 1, or 2) on each item obtained by any family member during individual testing. This composite score represents the highest score the group would have achieved had it consistently chosen as its collective response the best answer on each item proposed by any of its members during individual testing. The actual score is the score achieved by the group during its collective effort. Using these two scores, an efficiency ratio was calculated reflecting the extent to which the group in its collective effort utilized the resources of its individual members. The formula is:

$$\text{Efficiency Ratio} = 1 - \frac{\text{Maximum score} - \text{Actual score}}{\text{Maximum score}} \quad [1]$$

The efficiency ratio increases as the difference between the maximum and actual scores decreases.

The information regarding parent-child support was elicited from a systematic analysis of the group discussion adapted by Mills (1953) from Bales' (1950) process-analysis technique. Each verbal act is scored in sequence in one of three categories: contributions to the solution of the group problem, and supportive or nonsupportive reactions directed to others, classified on the basis of Bales' criteria. Scorable acts in this study were limited to verbal statements exclusively, and the basic unit for scoring was any statement expressing a complete thought. Supportive statements will hereafter be designated as S, nonsupportive statements as NS.

In order to equate individuals for total production of S and NS statements, ratios were used instead of absolute numbers of such statements. A support ratio was calculated for each parent in relation to the child in each of the two triadic groups. Similar ratios were calculated for patients in respect to each parent. The support ratio formula is simply:

$$\frac{S}{S + NS} \quad [2]$$

A high percentage indicates high support, a low percentage, low support.

The Wechsler responses and the group discussions were scored and coded blindly by two research assistants. Both coders, working independently, achieved 92% agreement on their coding of 10 pilot groups.

The measure of parental dominance was comprised of two parts: response dominance and total verbal activity. Response dominance occurs when one parent's response offered during individual testing appeared as the group's final response. If the group's answer matched the individual response of either parent alone, that parent was said to be dominant for that response. The total number of statements, as the second index of dominance, automatically results from the application of Bales' method which scores every statement in terms of speaker and listener (who-to-whom), as well as qualitatively.

RESULTS

Hypothesis I. Problem-solving efficiency

Prediction. The parent-patient triad will have a lower efficiency ratio than the parent-sibling triad when solving the problems from the Comprehension and Similarities subtests of the Wechsler-Bellevue Intelligence Scale.

A two-factor analysis of variance (with Groups of Families and Triads as the factors) applied to these efficiency ratios failed to yield any significant results. Each group of six families achieved a mean efficiency ratio of at least 85%, so that they fulfilled their problem-solving potential amazingly well. This high degree of efficiency does not conform to the reported descriptions of problem-solving "paralysis" in the families of schizophrenic patients.

The question may be posed whether this same degree of efficiency between the patient and sibling triads may not be an artifact produced by a compensatory maneuver of the parents who reject all of the patient's inadequate responses but accept the more adequate contributions of the sibling. In such an instance both groups could emerge as equally efficient. However, an analysis of the number of each child's individual responses which later appeared as the group's collective response failed to reveal any significant differences.

Hypothesis II. Parent-child support

Prediction. Parent-to-patient support ratios will be smaller than parent-to-sibling support ratios.

The difference between each parent's support ratios to the patient- and nonpatient-child was calculated. Neither the group of 24 fathers nor the group of 24 mothers showed a mean difference significantly different from zero. An inspection of the individual parent's support ratios to the patients and to their siblings revealed that many parents supported, or failed to support, both children in similar measure, while other parents did differ in their treatment of each child. However, the variability in the amount and direction of the differences was too great for any significant trend to emerge considering the size of the population. Thus, the evidence does not conform to the descriptions in the literature of marked differences in the interaction of parents with patients in contrast to the parents' treatment of their nonpatient children.

Hypothesis III. Parental role-reversal

Prediction A. Interparental reversals:

1. Mother's individual test responses should appear more often than father's as the

group response in the triad with male patients.

2. Father's individual test responses should appear more often than mother's as the group response in the triad with female patients.

3. Mother should produce a greater number of statements than father in the triad with male patients.

4. Father should produce a greater number of statements than mother in the triad with female patients.

5. Mothers' individual test responses should appear more often in triads with male patients than in triads with female patients.

6. Fathers' individual test responses should appear more often in triads with female patients than in triads with male patients.

The means and standard deviations of the number of parental responses appearing as the final group response to the Wechsler items are given in Table 1. In each family-patient triad, the difference between the response dominance of the parents was calculated by subtracting the number of dominant responses of the mother from those given by the father. Table 2 presents the mean differences for the families grouped according to the sex of the patient, as well as for all families disregarding the patient's sex. All these mean differences revealed that the father's individual responses appeared as the group response more often than did the mother's in all groups, regardless of the patient's sex. However, since none of the differences reached statistical significance, no evidence was found in support of Predictions A1 and A2 to the effect that parental response dominance would alter as a function of the patient's sex.

The number of parental statements made in the triads with the different sex patients was also analyzed with no significant findings. The mean number of statements made by the patients' fathers was consistently greater than the mean number of the mothers' statements, but the difference did not reach statistical significance. These "negative" findings also fail to support the predictions (A3 and A4) that the parents will be differentially active, verbally, as a function of the patient's sex. Both parents were equally active in terms of total speech production when in the presence of both male and female patients. Moreover, the tendency for fathers' statements to be more numerous than mothers' in both groups indicates that there was certainly no trend in the direction of less paternal than maternal participation in the group discussion.

TABLE 1

PARENTAL RESPONSE DOMINANCE IN PATIENT TRIADS

Group	N	Father		Mother	
		M	SD	M	SD
Male patients	12	5.42	4.52	3.08	3.03
Female patients	12	5.58	4.50	3.33	2.35

TABLE 2

DIFFERENCE BETWEEN PARENTAL RESPONSE DOMINANCE IN PATIENT TRIADS

Group	N	$M_{\text{Diff.}}$	$SD_{\text{Diff.}}$	t
Male patients	12	2.34	5.52	1.47
Female patients	12	2.25	6.55	1.19
All patients	24	2.30	5.92	1.90*

* $p < .10$.

Statistical analysis of the difference between the mean response dominance of the two groups of fathers (of male and female patients) and between the two groups of mothers failed to yield significant results. Therefore, no support was found for Predictions A5 and A6.

Prediction B. Parent-child sex-generation confusion:

1. Mother's support ratios to male patients should be larger than father's.

2. Father's support ratios to female patients should be larger than mother's.

3. Mother's support ratios to male patients should be larger than the support ratios of mother to female patients.

4. Father's support ratios to female patients should be larger than the support ratios of father to male patients.

5. Male patients will have larger support ratios to mothers than to fathers.

6. Female patients will have larger support ratios to fathers than to mothers.

Table 3 presents the mean and standard deviations of the parent-to-patient support ratios. Predictions B1 and B2 deal with the parental support to patients. In each family-

TABLE 3

PARENT-TO-PATIENT SUPPORT RATIOS

Group	N	Father			Mother		
		M	SD	t	M	SD	t
Male patients	12	29.67	14.09	2.15*	33.89	21.08	.39
Female patients	12	43.70	17.69		31.18	12.17	

* $p < .025$, one-tailed.

patient triad, the difference between the parents' support was calculated by subtracting the mother's support ratio from the father's. Table 4 gives the mean differences for families grouped according to the sex of the patient. With a significant mean difference of 12.52, the results support Prediction B2 to the effect that the father supported the female patient more than did the mother. Although the mother did support the male patient more than the father did, the −4.22 mean difference was not significant. Thus, the hypothesis of sex-generation confusion received only partial corroboration from the data gathered in this study.

Table 4 gives the comparisons for between-group differences (Subhypotheses B3 and B4). Prediction B4 is upheld since the group of fathers with female patients achieved a significantly higher mean support ratio than did the group with male patients. Again, no comparable finding emerged regarding the behavior of mothers toward their patient-sons in contrast to the support mothers offered their patient-daughters. Whatever crossing of generation boundaries occurred in the families was restricted to the father-daughter relationship, and was conspicuously absent in the mothers' relationships with their patient-sons.

The data which bear upon Predictions B5 and B6, to the effect that patients will support opposite-sex parents more than same-sex parents, did not reveal any significant findings. Patients did not support opposite-sex more than same-sex parents, and it is of interest to note that the female patients who received such marked support from their fathers did not reciprocate by evidencing any preferential support for them as compared to the support they offered their mothers.

TABLE 4

DIFFERENCE BETWEEN FATHER'S AND MOTHER'S SUPPORT RATIOS TO PATIENT

Group	N	$M_{\text{Diff.}}$	$SD_{\text{Diff.}}$	t
Male patients	12	−4.22	25.21	.58
Female patients	12	12.52	20.46	2.12*

* $p < .05$, one-tailed.

Of additional interest was an analysis of the interaction in one family dyad not dealt with in the hypotheses, namely, the child-to-parent behavior.

The means and standard deviations of the child-to-parent support ratios are presented in Table 5. Inspection of the mean support ratios indicated that the sex of the sibling was an important factor in determining the amount and direction of the difference between the patient and sibling support to each parent. Table 6 provides the results of the statistical analysis of the mean differences, with the families divided according to the sex of the sibling. For each parent, the sibling's support ratio was subtracted from the patient's support ratio. In those families with male siblings, the patients offered significantly

TABLE 5

CHILD-TO-PARENT SUPPORT RATIOS

Sex of patient		Patient triad		Sex of sibling		Sibling triad	
		Father	Mother			Father	Mother
Male	M	26.82	34.10	Female	M	29.90	22.25
	SD	16.60	22.79		SD	10.13	20.32
Male	M	33.82	40.70	Male	M	24.62	37.42
	SD	17.01	31.06		SD	16.33	12.29
Female	M	47.75	30.08	Male	M	16.18	22.67
	SD	15.70	17.71		SD	12.03	16.56
Female	M	36.58	47.48	Female	M	51.18	24.30
	SD	20.49	15.01		SD	17.30	12.46

Note.—24 patients; 24 siblings.

TABLE 6

DIFFERENCE BETWEEN PATIENT AND SIBLING SUPPORT RATIOS TO PARENTS' SUPPORT RATIOS

Families with	N	To father			To mother		
		$M_{\text{Diff.}}$	$SD_{\text{Diff.}}$	t	$M_{\text{Diff.}}$	$SD_{\text{Diff.}}$	t
Male siblings	12	20.40	18.15	3.89**	5.35	14.26	1.30
Female siblings	12	−8.84	27.98	1.09	17.52	25.54	2.38*

* $p < .05$.
** $p < .01$, two-tailed.

more support to their fathers than did their normal brothers, but the patients' support offered to their mothers was not significantly greater than that given by the nonpatient children. In those families with female siblings, the patients offered significantly more support to their mothers than did their normal sisters. The patients' sisters were more supportive of their fathers than were the patients, but this difference did not reach statistical significance. Further analysis of the child-to-parent support failed to reveal any evidence that the patients or male siblings tended to support an opposite-sex parent. However, the nonpatient females did express significantly greater support for their fathers than they did for their mothers: $t = 2.52$, $df = 11$, $p < .05$.

The final analysis concerns the difference between the number of S and NS statements exchanged among family members during the discussions. The intratriad statements were subdivided into their respective dyadic interactions: parent-to-child, child-to-parent, and interparental. The t tests revealed that in each case the NS statements were relatively more numerous, at the .01 level for the patient-to-parent dyad, at the .001 level for all other interactions, with patients and siblings alike. These findings convey some impression of the interfamilial dissension expressed during the group discussions but without proper control groups no estimate can be made regarding the extent to which the schizophrenic family is unique in this respect.

TABLE 7

DIFFERENCE BETWEEN NONSUPPORTIVE AND SUPPORTIVE STATEMENTS

Dyad	$M_{\text{Diff.}}$[a]	$SD_{\text{Diff.}}$	t[b]
Patient-to-parent versus sibling-to-parent	−17.83	34.70	2.52**
Interparental with patients versus interparental with siblings	14.71	35.73	2.02*

Note.—Mean difference of the differences.
[a] two-tailed tests, $df = 23$.
[b] NS > S statements.
* $p < .10$.
** $p < .05$.

Of special interest here are the differences between the quality of interaction of the various family members in the patient and sibling triads, such as parent-to-patient contrasted with parent-to-sibling; patient-to-parent contrasted with sibling-to-parents, etc. In each family the number of S statements exchanged between any two members was deducted from the number of NS statements. The difference found for each of the relationships within the sibling triad was then subtracted from the difference found in the parallel relationship within the patient triad in the same family.

A t of 2.52 ($p < .05$) reveals that the extent to which the NS statements predominated over the S statements was greater in the sibling-to-parent than in the patient-to-parent dyad. Also, the NS statements were relatively more numerous in the interparental dyad when the patient was present than when the nonpatient was the third party: the t of 2.02 approaches the .05 (2.07) level. Thus, the sibling was more critical of his parents than was the patient, while the interparental conflict was more pronounced in the presence of the patient- than in the presence of the nonpatient-child.

DISCUSSION

Contrary to the prediction, both parent-patient and parent-sibling triads exhibited an equal degree of efficiency in the solving of experimental problems. Similarly negative were the results relevant to the prediction that parents would address more S and fewer NS statements to their normal than to their patient-child. Parents were found to display great regularity in their support or nonsupport of both children. The findings concerning the presence of parental role-reversal also failed to substantiate predictions. Although the differences were not statistically significant, fathers' responses predominated over mothers' in groups with both male and female patients, and no reversal of roles was evident. In terms of verbal productivity, both parents were equally active, and no evidence was found with this measure to support the hypothesis of the aloof, passive father. These negative findings concur with reports from other behavioral experiments with families which also failed to corroborate the role-reversal theory (Caputo, 1963; Cheek, 1964; Farina & Dunham, 1963). In a study of psychological test protocols, the fathers of schizophrenic children were described as "driving" personalities (Block, Patterson, Block, & Jackson, 1958). The accumulating evidence thus suggests that there need not be any breakdown of the traditional parental roles associated with the emergence of severe pathology in a family member.

The second part of the role-reversal hypothesis predicts greater mutual support between patients and opposite-sex parents than between patients and same-sex parents. The results offered partial corroboration of the hypothesis. Fathers were more supportive of their patient-daughters than were patient-daughters' mothers, and these fathers were even more supportive of their daughters than were other fathers of their patient-sons. There was, therefore, a particularly strong attachment between fathers and female patients, but no similar trend was apparent in the mother-son relationship. This is contrary to the hypothesis as well as to the findings reported by Lennard et al. (1965), though it is consistent with the experimental findings reported by other investigators, as noted above.

In contrast to the regularity manifested in the parents' S statements addressed to both children, the children, on their part, evidenced distinct differences in the amount of support they addressed to their parents. Both children behaved similarly toward one parent, but differently toward the other parent. The normal children were far more critical of a same-sex parent than were the patients toward the identical parent. Both children were equally critical or supportive of the parent who was of the opposite sex of the sibling. Thus, the parent of the same sex as the normal child became the focus of divergent allegiances on the part of the children. In this manner the family may become divided into rival dyads, as suggested by some observers (Lidz, 1963). However, no evidence was found for the contention that patients tend to support an opposite-sex parent more than their same-sex parent. Even though the fathers exhibited marked support of their patient-daughters in contrast to that offered these girls by their mothers, the female patients did not reciprocate by expressing preferential support for their fathers over their mothers. Remarkably enough, it was the normal daughter who displayed greater support for her father than for her mother.

An analysis of the number of S and NS statements exchanged by family members revealed that, as with the ratios, parents did not vary in the relative number of such statements they addressed to both children. However, of great interest was the finding that the relative number of S to NS statements did differ in the child-to-parent and in the interparental dyads as a function of which child was present in the group. The patients were more supportive of their parents than were the siblings, and there was a corresponding change in the interparental support in which the parental discord was more marked in the presence of the less critical patient than it was in the presence of the more critical sibling. In this sense, the patient was associated with greater interparental conflict than was the normal child while he appeared to be more directly involved in conflict with the parents, or with one parent at least, than was the patient. This finding highlights the importance of the interparental relationship in the study of the patients' families, as is emphasized by Lidz (1963) and others.

Consequently, these families manifested a shift in the balance of their various dyadic interactions as a function of the child's patient or nonpatient status. The behavior of the family members was interrelated in what may be characterized as a system-inequilibrium, in which the children expressed varying degrees of criticism toward their parents, while the parents' expression of mutual criticism altered as a function of the child's behavior. Whether the more passive or active posture of the patient and sibling was a cause or effect of the parental posture, and whether this parameter was relevant to schizophrenia, cannot be established from these data.

There was a confluence of findings pointing to mother's generally unsupportive behavior in these families. As noted, she was less supportive of her patient-daughter than was the father. It was learned, too, that the normal daughter supported her father more than she supported her mother, so that she may have found her mother to be critical and unaccepting. Cheek (1964) described the mothers in her study as reminiscent in many ways of the so-called schizophrenogenic mother. The hypothesis suggested here, however, is not intended as a reversion to the theory of the schizophrenogenic mother. Rather, it emphasizes how a rejecting mother's attitude influences the emotional atmosphere in the entire family.

The measures employed in this study were subject to a number of limitations, the most salient of which was, perhaps, the highly structured nature of the task and its relationship to intelligence factors. Furthermore, the number of statements expressed during a discussion may not provide a completely reliable index of parental dominance since a verbally productive parent may still be submissive in other ways. These findings point to the need to delineate more precisely the relevant parameters of what is globally referred to as a dominant or submissive parent. It is also possible that the behavioral experiment, with its cross-sectional framework, was too artificial to assess the subtle nuances of enduring family relationships which could play an important role in the emergence of mental illness (Handel, 1965).

REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.

BATESON, G. Minimal requirements for a theory of schizophrenia. *Archives of General Psychiatry,* 1960, 2, 477-491.

BLOCK, J., PATTERSON, V., BLOCK, J., & JACKSON, D. S. A study of schizophrenic and neurotic children. *Psychiatry,* 1958, **21**, 387–397.

BOWEN, M. A family concept of schizophrenia. In D. D. Jackson (Ed.), *The etiology of schizophrenia.* New York: Basic Books, 1960. Pp. 346–372.

BRODEY, W. M. Some family operations of schizophrenia: A study of five hospitalized families each with a schizophrenic member. *Archives of General Psychiatry,* 1959, **1**, 379–402.

CAPUTO, D. V. The parents of the schizophrenic. *Family Process,* 1963, 2, 339–356.

CHEEK, F. The schizophrenogenic mother in word and deed. *Family Process,* 1964, **3**, 155–177.

FARINA, A., & DUNHAM, R. Measurement of family relationships and their effects. *Archives of General Psychiatry,* 1963, **9**, 64–73.

FLECK, S., LIDZ, T., & CORNELISON, A. Comparison of parent-child relationships of male and female schizophrenic patients. *Archives of General Psychiatry,* 1963, **8**, 1–7.

HALEY, J. The family of the schizophrenic: A model system. *Journal of Nervous & Mental Disorders,* 1959, **129**, 357–374.

HALEY, J. Family experiments: A new type of experimentation. *Family Process,* 1962, **3**, 41–65.

HANDEL, G. Psychological study of whole families. *Psychological Bulletin,* 1965, **63**, 19–41.

LENNARD, H. L., BEAULIEU, M. R., & EMBREY, N. G. Interaction in families with a schizophrenic child. *Archives of General Psychiatry,* 1965, **12**, 166–183.

LIDZ, T. *The family and human adaptation.* New York: International Universities Press, 1963.

LIDZ, T., CORNELISON, A., FLECK, S., & TERRY, D. The intrafamilial environment of the schizophrenic patient: IV. Parental personalities and family interaction. *American Journal of Orthopsychiatry,* 1956, **28**, 764–776.

LIDZ, T., FLECK, S., ALANEN, Y., & CORNELISON, A. Schizophrenic patients and their siblings. *Psychiatry,* 1963, **26**, 1–18.

LU, YI-CHUANG. Mother-child role relations in schizophrenia. *Psychiatry,* 1961, **24**, 133–142.

LU, YI-CHUANG. Contradictory parental expectations in schizophrenia. *Archives of General Psychiatry,* 1962, **6**, 219–234.

MILLS, T. Power relations in three-person groups. *American Sociological Review,* 1953, **18**, 351–357.

PROUT, C. T., & WHITE, M. A. The schizophrenic sibling. *Journal of Nervous & Mental Disorders,* 1956, **123**, 162–170.

ROMAN, M., & BAUMAN, G. Interaction testing: A technique for the psychological evaluation of small groups. In M. Harrower (Ed.), *Creative variations in the projective techniques.* Springfield, Ill.: Charles C Thomas, 1960. Pp. 93–138.

WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

WECHSLER, D. *The Wechsler-Bellevue Intelligence Scale.* Form II. New York: Psychological Corporation, 1946.

(Received July 9, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 354–362

EFFECTS OF SOCIAL REINFORCEMENT ON THE LEVEL OF ABSTRACTION IN SCHIZOPHRENICS[1]

DONALD H. MEICHENBAUM[2]

University of Illinois

This study was designed to investigate (a) the effects of the nature of contingent vs. noncontingent social reinforcement upon the level of abstraction in schizophrenics as measured on a proverbs task, (b) the generalization of this behavior to another verbal conceptual task administered under neutral conditions, (c) the predictive and construct validity of the Ullmann-Giovannoni Process-Reactive Self-Report Scale. 64 male schizophrenics were assigned to 4 treatment groups: contingent positive, noncontingent positive, contingent negative, and control. The results indicated that only the contingent positive group significantly increased their level of performance on the proverbs task and showed generalization. The other 3 groups showed no significant differences. Validity evidence for the Process-Reactive Scale and marital status was obtained.

A significant difference between the conceptual behavior of schizophrenics and normals is well established. Buss and Lang (1965) have reviewed several conceptualizations of the schizophrenic thought disorder. One approach considers schizophrenia as a disorder which involves a relatively permanent loss of conceptual ability. Investigators such as Goldstein (1959), Hanfmann and Kasanin (1942), and Rapaport (1945) hold this view. In many investigations cited in support of this approach, adequate control groups were absent and inadequate statistical procedures were used. Buss and Lang indicated several later investigations (Chapman & Taylor, 1957; Lothrop, 1960; White, 1949) which showed that schizophrenics are capable of responding with abstract concepts.

An alternative approach toward schizophrenic conceptual deficit views the problem as a lack of communication. Sullivan (1946) and Cameron (1946) emphasized that the schizophrenic disorder involves a general disturbance in social communication and a motivational problem arising from an overall fear of interpersonal relationships. Cameron held that some abstract ability can be demonstrated in schizophrenics if the examiner concentrates on establishing rapport, eliciting cooperation, and fostering a total positive milieu. The idiosyncratic, eccentric, and private meanings of concepts in schizophrenic thought have been indicated by McGaughran and Moran (1956) and Payne and Hewlett (1960). Buss and Lang (1965), in reviewing this literature, suggested that both the dimension of concreteness and privacy may be confounded with a more general intellectual disability. They also emphasized that the incomprehensibility of the schizophrenic's concepts was due to the bizarre nature of the concepts rather than an inability to communicate them.

Lothrop (1961) reviewed the literature on schizophrenic cognitive deficit and indicated that the level of conceptual performance of schizophrenics varies as a function of the content and structure of the conceptual task, the nature of the interpersonal situation, and the composition of the schizophrenic population. The present investigation concerns the effects of the interpersonal nature of the test situation on the level of conceptual ability in schizophrenics.

Two major experimental designs have been used to manipulate the interpersonal nature of the test situation. The first follows the paradigm of verbal operant conditioning in which the experimenter (E) responds "contingently" to some specified class of a sub-

---

[1] The author wishes to thank the staff of the Veterans Administration Hospital, Danville, Illinois, particularly Angela Folsom and Roy Hamlin for their help in the collection of data; and Leonard P. Ullmann and Wesley C. Becker for their constructive suggestions in planning this research and their critical reading of the manuscript.

[2] Now at the University of Waterloo, Waterloo, Ontario, Canada.

ject's (*S*'s) responses. Many important and primary symptoms in schizophrenia have been subjected to manipulation by this method in both laboratory and clinical settings. Salzinger and Pisoni (1958) demonstrated that affect self-references could be influenced by verbal conditioning. Weiss, Krasner, and Ullmann (1963), Ullmann, Krasner, and Edinger (1964), and Ullmann, Forsman, Kenny, McInnis, Unikel, and Zeisset (1965), using verbal operant techniques, demonstrated control of the amount and level of emotional words, common associations, and "sick talk" elicited from schizophrenics. Ayllon and Haughton (1964), in a clinical setting, used verbal conditioning by ward personnel over a period of several months to systematically manipulate bizarre verbal behavior.

The second experimental approach varies the total milieu between *E* and *S* in order to test the motivational hypothesis that schizophrenic behavior represents a defensive maneuver on the part of the patient when placed under a condition of interpersonal stress. The *E*'s behavior is "noncontingent" upon any class of an *S*'s responses. Schupper (1955), assuming an accepting and rejecting role, found that under the rejecting condition chronic schizophrenics displayed a generally poor performance on a variety of tasks, whereas acute schizophrenics were defective in concept formation. Hirschman (1954) and Feldstein (1962) found that the nature of the *E*'s interpersonal involvement did not differentially affect schizophrenics' verbal behaviors. Kaufman (1963) manipulated the relative amount of positive and negative social reinforcement provided during a standard interview between administrations of parallel forms of a Proverbs Test. She found that the positive group showed a marked improvement in their performance, while the negative group significantly decreased in their performance. Kaufman (1960, 1963) found that the parallel forms of the Proverbs Test which she developed by means of an item analysis were sensitive and reliable measures of abstract-thinking pathology in schizophrenia. The items of the two forms maximized the correlation with the criterion measured (Modified Elgin Prognostic Scale; Becker, 1956) and minimized the correlation with a standard vocabulary test. Kaufman's (1960) Parallel Proverbs Test was used in the present investigation.

A principal limitation of research in this area has been the heterogeneity of *S*s diagnosed as schizophrenic. Two scales which have been used to assess the degree of pathology are the Modified Elgin Scale and the Phillips Scale (Phillips, 1953). Both scales are time consuming, require a great deal of effort, and are subject to extraneous sources of variance, for example, the varied quality of social and medical records. These limitations were overcome by Ullmann and Giovannoni (1964) who developed a 24-item, true-false self-report scale on the basis of internal consistency which obtains biographical information to reflect the nature of the patient's prior social interaction with his environment. This process-reactive self-report scale was used in the present study.

Until recently, few investigators were concerned with the problem of transfer or generalization of verbal conditioning effects from one stimulus situation to another. Drennen (1963) reviewed this work and reported that five of nine experiments obtained results which were considered statistically significant regarding the generalization of the effects of verbal conditioning. He demonstrated that the generalization of conditioning was a function of the variable of task-cue similarity. Ullmann, Krasner, and Collins (1961) and Ullmann, Krasner, and Ekman (1961) showed that patients reinforced for using emotional words in an experimental situation showed improvement in a group-therapy situation; patients not so reinforced showed no gain. The present experiment evaluated the extent of response generalization.

The purposes of the present investigation were:

1. To examine the effects of the nature of contingent versus noncontingent social reinforcement on the level of abstract conceptual performance of schizophrenics as measured by a Proverbs Test.

2. To examine the generalization of this behavior to another verbal conceptual task administered under neutral conditions; specifically, the Similarities subtest of the

Wechsler-Bellevue Adult Intelligence Scale (WAIS), Form II (1946).

3. To examine the predictive and construct validity of the Ullmann-Giovannoni Process-Reactive Self-Report Scale.

METHOD

Subjects

The *Ss* were 64 male schizophrenics from the Veterans Administration Hospital, Danville, Illinois, all less than 50 years old, and with no indication of central nervous system pathology; *Ss* were selected from acute admission wards and back wards. All were under medication. Average age of the sample was 38.51 years, with a range of 24–48; average length of present hospitalization was 13.17 months, with a range of 1–48 months. The 64 *Ss* were randomly assigned to four groups of 16 *Ss* each—three were experimental and one was a control group. Within each group, 8 were classified as process (scores 13–24) and 8 as reactive (scores 0–12) according to their scores on the Ullmann-Giovannoni (1964) Process-Reactive Self-Report Scale.[3]

Procedure

Each *S* was seen for approximately 1 hour on 2 consecutive days and told that he had been selected to help in preparing a new test to be used in the hospital. On Day 1, each *S* was given the following sequence of tests all scored according to standard criteria: the Process-Reactive Self-Report Scale; the Vocabulary subtest of the 1960 Stanford-Binet, Form L-M; the Similarities subtest of the 1939 WAIS, Form I. The following day, each *S* was given in counterbalanced order the Kaufman (1960) Proverbs Test consisting of parallel forms of 12 items each. There were 24 additional proverbs chosen from the *Dictionary of American Proverbs* (Kin, 1955) and from other tests which were administered to each *S* between the parallel forms of the Kaufman Proverbs Test.

The following instructions and warm-up example were given to each *S* prior to the Proverbs Test:

> Today Mr. .... I want to ask you to explain the meaning of some proverbs as best you can. Do you know what a proverb is? (Pause for *S*'s answer.) That's right, (if appropriate) a proverb is a saying which has a general meaning, or an idea which goes beyond the words in the saying itself. Remember, you are to tell what the proverb means rather than just telling more about it.
>
> Now, I want you to give me the meaning or idea behind each proverb or saying. Let's try one and see if you can get the meaning.
>
> Example: We only know the worth of water when the well is dry. (Pause for *S*'s answer.) That's right, (if appropriate) the proverb has a general meaning, namely, we don't appreciate the value of things until we are deprived of them, or we don't know when we're well off until we're poor.
>
> Now let's try some other proverbs.

Next the Similarities subtest of the 1946 WAIS, Form II, was administered under neutral standard testing conditions to all *Ss*. This test was included to assess the extent of generalization. Each *S* was then seen by an advanced clinical psychological trainee who administered a structured interview which assessed *S*'s perception of *E*'s attitude and behavior. The interview measured the degree to which the experimental conditions were achieved.

A tape recorder, inconspicuously placed in a corner of the room, recorded both proverb testing and the subsequent structured interview, thus permitting the examiner to respond immediately and naturally. All proverb responses on Kaufman's parallel forms were keyed and transcribed to 3 × 5 inch index cards and submitted to two independent raters. To minimize serial position and halo effects, each proverb was rated across *Ss* before rating the next proverb. Although no rater reliability was determined for the additional 24 proverbs, each protocol was keyed at the time of testing and scored several months later across *Ss* to reduce the likelihood of bias. The structured interviews were similarly transcribed, keyed, and submitted to three independent raters.

Each proverb was rated for level of abstraction according to the criteria established by Becker (1956). The categories of classification are: Absurd (weight of 1), Literal (weight of 2), General-Literal (weight of 2), False Interpretation (weight of 2), Vague Response (weight of 3), and Abstract I, II, and III, respectively weighted 4, 5, and 6.

Treatments

The first 12 of the total 48 proverbs administered to each *S* represented an operant measure; the last 12 represented a posttreatment test. The relative differences in *Ss*' performances on the parallel proverbs forms reflect the differential effects of the various experimental conditions. The following treatment conditions were applied to the last 36 proverbs.

Group I—Contingent Positive Group. The *Ss* comprising this group were positively reinforced for abstract responses to proverbs, namely, proverb responses of the quality of Abstract III, II, and I scored 6, 5, and 4, respectively, in Becker's (1956) scoring system. Following the initial operant period, any abstract response to the remaining 36 proverbs was reinforced with "mmh-hmm," smile, head nod, followed by a positive encouraging "good." The *E*'s manner was positive, helpful, interested, and warm only when abstract responses were offered on the last 36 proverbs.

---

[3] The scale was administered verbally to each *S* because pilot work indicated that patients often misinterpreted the questions. Question 15 was changed to read, "More than once in the last year after some group meeting I have talked with some other members about something that went on." The phrase "I have stayed on" was deleted because the patients in group meetings at VA Danville are typically ushered out of the meeting room at the completion of the group session.

Group II—Noncontingent Positive Group. The reinforcement schedule for *Ss* in this group was selected to match *S*-for-*S* the frequency and schedule of positive reinforcements administered to an *S* in the reinforcement group. Thus, an *S* in Group II would receive exactly the same pattern and type of reinforcement as the matched *S* in Group I, regardless of the level of interpretation of *S*'s proverb response. The inclusion of Group II provides an evaluation of the contingency variable versus the total milieu effect of positive reinforcement on the level of conceptual ability in schizophrenics.

Group III—Contingent Negative Group. The *Ss* comprising this group were negatively reinforced for nonabstract responses to proverbs, namely, proverb responses of the quality vague, literal, general-literal, false interpretation, and absurd, scored 3, 2, and 1. Following the initial operant period, any nonabstract response to the remaining 36 proverbs was reinforced with "uh-uh," frown, shaking of the head "no," followed by a negative, discouraging "poor." The *E*'s manner was negative, rejecting, and cold only when nonabstract responses were offered on the last 36 proverbs.

Group IV—Control Group. The *Ss* in this group received no reinforcement. A constant neutral standard test attitude was maintained by *E* for the entire test administration. No marked smiling or frowning and no words of encouragement or disapproval were offered. All conversation and actions were completely task oriented.

An accuracy tally or "hit rate" (percentage of proverb responses correctly reinforced) and a "false alarm rate" (percentage of proverb responses reinforced when they should not have been reinforced) were determined for both the contingent positive and contingent negative groups.

RESULTS

Equivalence of Groups

Equivalence of the four treatment groups in terms of demographic variables (years of education, marital status, length of present hospitalization) and Binet Vocabulary was obtained.[4] The contingent positive group was significantly younger than contingent negative and noncontingent positive groups. However, since age did not correlate significantly with the dependent measures of proverb change scores or similarities change scores, this factor was not given further consideration. An analysis of variance of process-reactive scores indicated that the groups did not differ significantly on this self-reported indicator of severity of disorder ($F = .344$, $df = 3/60$, $p > .05$). An analysis of variance of the first administration of both the proverb scores and the Similarities subtest scores yielded $F = .266$ ($df = 3/60$, $p > .05$), and $F = 1.41$ ($df = 3/60$, $p > .05$), respectively. These results indicated that the four treatment groups were equivalent with respect to performance on the measures of abstract verbal ability prior to differential treatment.

Table 1 presents the means and standard deviations of demographic and experimental variables before experimental manipulation for 32 reactive and 32 process *Ss*. The process group, as compared to the reactive group, obtained significantly inferior scores on the Proverbs Test ($t = 2.81$, $df = 62$, $p < .01$) and had significantly longer length of present hospitalization ($t = 3.12$, $df = 62$, $p < .01$). Other comparisons were not significant.

Before examining the effects of the experimental manipulation on the level of abstract verbal performance, the degree to which the differential experimental treatments were achieved was evaluated. The "hit rate" and the "false alarm rate" for the contingent positive group were 92% and 10%, respectively; for the contingent negative group 95% and

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR SUBJECTS CLASSIFIED REACTIVE AND PROCESS ON THE ULLMANN-GIOVANNONI SCALE BEFORE EXPERIMENTAL MANIPULATION

Variable	Reactive		Process	
	M	*SD*	*M*	*SD*
Process-reactive score	7.19	2.99	16.78	2.51
Age	38.78	5.61	38.25	5.80
Years of education	12.53	3.69	11.53	2.30
Length of present hospitalization (mos.)	8.09	8.30	18.25	16.45
Stanford-Binet Vocabulary score	24.44	2.05	22.84	5.54
WAIS Similarities score	12.03	4.03	10.78	3.85
Proverbs score	37.23	8.64	31.44	7.82

---

[4] The contingent positive group showed a slightly higher but not significant mean for years of education and Binet Vocabulary scores; and a somewhat lower mean for length of present hospitalization. However, these three variables did not correlate significantly with the change scores on the Proverbs Test or Similarities subtest for either the total sample or for the four respective treatment groups. These correlations provided further evidence for the initial equivalence of the treatment groups.

TABLE 2

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR PARALLEL PROVERB SCALES

Source	*df*	*MS*	*F*
Experimental group	3	323.06	9.62*
Process-Reactive	1	8.62	.257
Group × Process-Reactive	3	9.92	.296
Within cells	56	33.57	

* $p < .001$.

15%, respectively. These data indicated that *S*s' proverb responses were correctly and reliably reinforced by *E*.

Three advanced clinical students sorted the structured interviews on the basis of whether *S* had been exposed to a positive, negative, or neutral experimental condition. The contingent positive and noncontingent positive groups were combined for this evaluation. The judges agreed on 75% of all their assignments to the respective categories, indicating that the correspondence between *E*'s intended effect and *S*'s perception of the experimental conditions had been achieved.

A correlation coefficient of .950 ($N = 64$) was obtained between two independent raters for the parallel forms of the Proverbs Test. The nonsignificant mean difference ($t = 1.52$, $df = 62$, $p > .10$) indicated the comparability of the parallel forms of the Proverbs Test.

Effects of Experimental Conditions

As shown in Table 2, an analysis of variance of the differences between scores on the proverb scales for the four treatment groups yielded a significant effect for experimental conditions. Multiple comparison of group means was performed using Duncan's test (Edwards, 1962), which indicated that the contingent positive group was significantly different ($p < .05$) from the noncontingent positive, contingent negative, and control groups. The latter three groups did not differ significantly from each other. The mean change score for *S*s in the contingent positive group was 8.28; for those in the noncontingent positive group, .69; for those in the contingent negative group, −.22; and for those in the control group, −1.84.

In order to determine the differential effects of the treatment conditions during acquisition performance on the proverbs task, a 4 × 2 × 9 analysis of variance (four experimental groups by process-reactive by nine experimental blocks of trials) was performed. As shown in Table 3, the main effect of experimental group was significant, with the contingent positive group performing superior to the noncontingent positive, contingent negative, and control groups. The latter three groups did not significantly differ from each other.

The significant *F* obtained for the main effect of level of reported disorder (process versus reactive) indicated that the reactive group's performance on the proverb scales was significantly superior to that of the process group. This difference, as well as the significant initial difference on the operant trials, indicated that the Proverbs Test was a sensitive index of level of thought disorder and was reliably related to level of self-reported disorder. Neither the interactions nor the trial effects were significant.

Figure 1 shows the four treatment groups' performances on the proverbs task in 12 blocks of trials of four proverbs each. Each point represented the average of the total responses of 16 *S*s to four proverbs. The range of scores is 4–24. The experimental manipulation was imposed on the last 36 proverbs, that is, Blocks 4–12. The superior performance of the contingent positive group was clearly indicated.

An analysis of variance was performed on the differences between scores on the parallel

TABLE 3

ANALYSIS OF VARIANCE OF ACQUISITION PERFORMANCE ON THE LAST 36 PROVERBS

Source	*df*	*MS*	*F*
Between subjects	63		
Experimental group	3	197.15	4.18*
Process-Reactive	1	547.36	11.62*
Group × Process-Reactive	3	17.83	.378
Error	56	47.11	
Within subjects	512		
Trials	8	19.81	.743
Group × Trials	24	9.59	.359
Process-Reactive × Trials	8	5.59	.209
Group × Process-Reactive × Trials	24	4.61	.173
Error	448	26.66	

* $p < .01$.

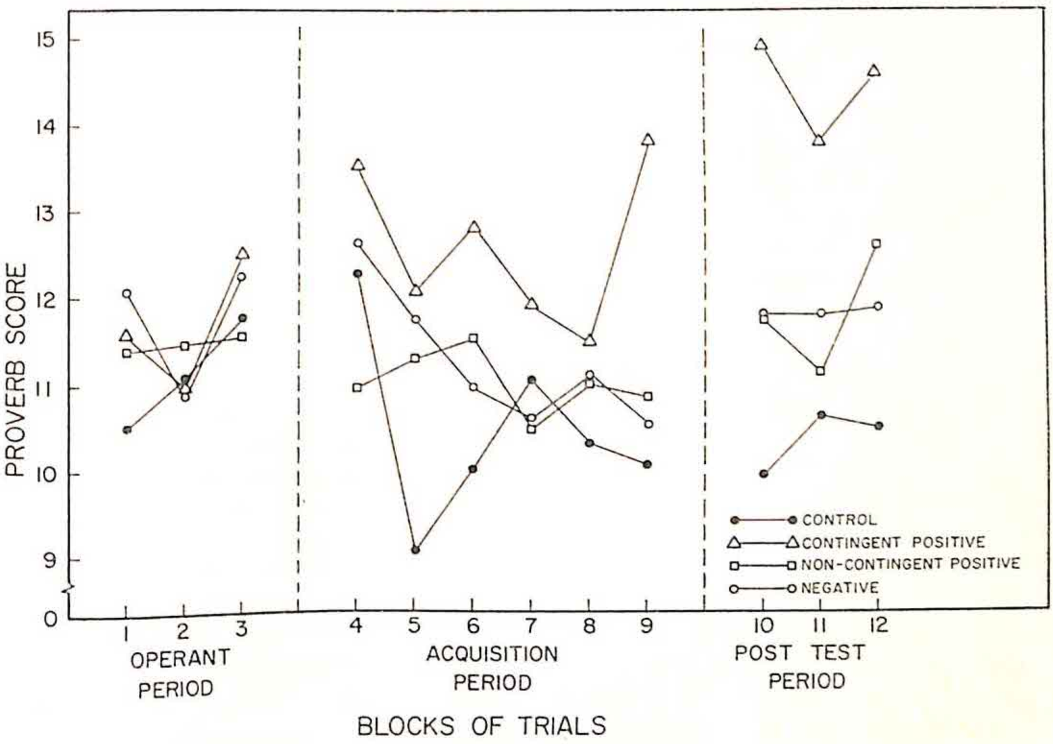


FIG. 1. Proverb performance of four treatment groups on operant, acquisition, and posttest period.

Similarities subtests (WAIS, Forms I and II) in order to determine the extent of generalization of "abstracting ability" to another verbal conceptual task which was administered under neutral conditions. As shown in Table 4, a significant effect for experimental conditions was obtained. Tests of multiple comparison indicated that the contingent positive group was significantly different ($p < .05$) from the noncontingent positive, contingent negative, and control groups. The latter three groups did not differ significantly from each other. The mean change score for *S*s in the contingent positive group was 2.19; for those in the noncontingent positive group, − .50; for those in the contingent negative group, − .69; and for those in the control group, .06. Thus, only the contingent positive group showed significant improvement on the proverbs task and on the similarities generalization task.

TABLE 4

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR PARALLEL SIMILARITIES SUBTESTS

Source	*df*	*MS*	*F*
Experimental group	3	27.89	5.29*
Process-Reactive	1	.39	.074
Group × Process-Reactive	3	2.43	.461
Within cells	56	5.27	

* $p < .01$.

Validation of the Process-Reactive Self-Report Scale

Predictive Validity. The criterion of predictive validity employed was the total number of days (TDI) the patient spent in the hospital from the day the study was completed until follow-up information was obtained. The discharge of the patient from the hospital was entirely independent of his score in the Process-Reactive Self-Report Scale. The maximum number of days a patient could spend in the hospital was 170 and the minimum was 1. For example, if a patient had received a leave of absence of 10 days and a trial visit of 60 days, his TDI score would have been 100 days. Because the sampling distribution of TDI was markedly skewed with the largest number of *S*s remaining in the hospital 170 days, 3 × 3 class contingency coefficients (*C*) were computed. The relationship between TDI and process-reactive scores yielded a *C* of .455 ($\chi^2 = 16.47$, $df = 4$, $p < .01$). This result indicated that those *S*s who scored toward the reactive end of the scale were discharged sooner; whereas those *S*s who scored toward the process end of the scale remained in the hospital.

One of the major classes of items in many prognostic scales, such as the Phillips Scale or the Modified Elgin Scale, concerns the heterosexual behavior and marital status of *S*. The separate contributions of the process-reactive scale less the four marital status items, and a 3-point marital status scale (1-married, 2-divorced, and 3-single) in predicting the TDI criterion were examined. The *C* between process-reactive scale less four marital items and TDI was .425 ($\chi^2 = 13.88$, $df = 4$, $p < .01$). This finding indicated that the remaining items on the process-reactive scale contributed to the predictive value of the scale. The *C* between the 3-point marital status scale and TDI was .442 ($\chi^2 = 13.68$, $df = 4$, $p < .01$). Married patients left the hospital sooner, whereas single patients remained in the hospital longer.

Neither the full-scale process-reactive score nor the scores on the simple 3-point scale correlated significantly with the type of discharge the patient received, for example, maximum hospital benefits, trial visit, and leave of absence.

Construct Validity. One kind of evidence for the validity of prognostic indicators is a significant negative relationship with level of thinking pathology in schizophrenics, namely, those *S*s who indicate a more adequate level of prior social adjustment do better on conceptual tasks, whereas those *S*s who evidence less adequacy do more poorly on conceptual tasks. As shown in Table 5, significant negative correlations were obtained between total scores on the verbal conceptual tasks administered prior to the differential experimental treatments and total scores on the process-reactive scale. Table 5 also presents the set of intercorrelations between the three conceptual tasks. Contrary to findings by Becker (1956) and in accord with Kaufman's (1963) findings, correction for verbal intelligence does not increase the correlation between proverbs and process-reactive scores.

TABLE 5

CORRELATIONS BETWEEN CONCEPTUAL TASKS AND PROCESS-REACTIVE SCORES

Task	Proverbs	Vocabulary	Similarities	Process-Reactive
Proverbs	—	.637**	.553**	−.281*
Vocabulary		—	.661**	−.234*
Similarities			—	−.247*

Note.—$N = 64$.
* $p < .05$.
** $p < .001$.

The construct validity of the 3-point marital status scale was also determined. The *C* between the 3-point scale of marital status and proverbs scores was .392 ($\chi^2 = 11.73$, $df = 4$, $p < .02$); between marital status and vocabulary scores, .338 ($\chi^2 = 8.78$, $df = 4$, $p < .10$); and between marital status and similarities, .115 ($\chi^2 = 5.26$, $df = 4$, $p < .30$). These results indicated that those schizophrenics who were married performed more adequately on the proverbs task; whereas schizophrenics who were single performed less adequately. A similar trend was evidenced on the vocabulary task.

Proverbs, vocabulary, and similarities scores were not significantly correlated with TDI.

DISCUSSION

Analyses of the proverb change scores and the acquisition performance of the four treatment groups indicated that the contingent positive group differed significantly from the noncontingent positive, contingent negative, and control groups. The latter three did not differ significantly from each other. The importance of the contingency parameter of reinforcement was clearly indicated.

The lack of improvement in the contingent negative group may be due to two noticeable patient reactions evidenced during testing and upon administration of the structured interview. One type of reaction involved emotional and hostile responses elicited by negative reinforcement. Eight *S*s in this group decreased the level of their performances on the Proverbs Test (negative change scores), two *S*s remained unchanged, and six *S*s improved the level of their performances (positive change scores). Hostile and emotional responses were evidenced by six of the eight *S*s who decreased the level of their performances and by one *S* whose performance remained unchanged. The emotional and hostile responses appeared to interfere with *Ss*' cognitive functioning and to cause a decrement in performance. The other reaction was marked by *Ss*' attempts to use the negative reinforcement as information, thus improving their performances. The *Ss*' behavior in the latter group did not appear to differ from *Ss*' behavior in the contingent positive group. The validity of this distinction is underlined by the significant negative Spearman rank correlations for the negative reinforcement group between proverb change scores and TDI (rho $= -.519$, $p < .05$), and proverb change scores and process-reactive scores (rho $= -.786$, $p < .01$). These results indicated that proverb change scores for the negative group were significant prognostic indicators; reactive schizophrenics improved their performances significantly only in the negative group while process schizophrenics decreased their performance level. Further research is planned to investigate the contingent versus noncontingent effects of negative reinforcement upon schizophrenic performance and its interaction with process-reactive continuum.

The generalization of "abstracting ability" to another verbal conceptual task administered under neutral conditions was also examined. The analysis of change scores on the WAIS Similarities subtest (Form I versus Form II) indicated that the contingent positive group showed significant generalization of improved performance. The other three treatment groups did not differ significantly from each other. Many interesting research possibilities are raised by including a generalization task within an experimental design. For example, will the patient generalize with another *E*? How distant can the generalization task be removed from the conditioning task in content and time and still obtain evidence of generalization?

The final aim of this investigation was to examine both the construct and predictive validity of the Ullmann-Giovannoni Process-Reactive Self-Report Scale. Significant negative correlations were found between the level of thinking pathology and level of prior social adjustment. A similar significant relationship was found between marital status and Proverbs Test scores. Predictive validity for both the process-reactive scale and marital status was obtained. Farina, Garmezy, and Barry (1963) have reviewed several studies of prognosis in schizophrenia and conclude that marital status is as good an indicator of probability of discharge from the hospital as more complex scales. The present findings confirmed their conclusions.

In conclusion, the findings of this investigation have indicated that the nature of the conceptual deficit in schizophrenia is inadequately explained as a relatively permanent loss in conceptual ability since the contingent positive group showed a marked improvement in conceptual ability. Furthermore, the failure of the noncontingent positive group to improve indicated that a total positive milieu is not a sufficient condition for reducing conceptual deficit. Thus, the contingency of reinforcement is a crucial variable in reducing conceptual deficit in schizophrenics. The contingency of reinforcement provides informational or feedback cues which indicate adequacy of performance. Further research is necessary to determine the extent to which conceptual deficit in schizophrenia can be modified by use of contingent reinforcement over longer periods of time.

REFERENCES

AYLLON, T., & HAUGHTON, E. Modification of symptomatic verbal behavior of mental patients. *Behaviour Research and Therapy*, 1964, **2**, 87–99.

BECKER, W. C. The relation of severity of thinking disorder to the process-reactive concept of schizophrenia. Unpublished doctoral dissertation, Stanford University, 1956.

BUSS, A. H., & LANG, P. J. Psychological deficit in schizophrenia: I. Affect, reinforcement, and concept attainment. *Journal of Abnormal Psychology*, 1965, **1**, 2–24.

CAMERON, N. S. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: University of California Press, 1946. Pp. 50–64.

CHAPMAN, L. J., & TAYLOR, J. A. The breadth of deviate concepts used by schizophrenics. *Journal of Abnormal and Social Psychology*, 1957, **54**, 118–123.

DRENNEN, W. T. Transfer of the effects of verbal conditioning. *Journal of Abnormal and Social Psychology*, 1963, **66**, 619–622.

Edwards, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1962.

Farina, A., Garmezy, N., & Barry, H. Relationship of marital status to incidence and prognosis of schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **67**, 624–630.

Feldstein, S. The relationship of interpersonal involvement and affectiveness of content to the verbal communication of schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1962, **64**, 39–45.

Goldstein, K. Concerning the concreteness in schizophrenia. *Journal of Abnormal and Social Psychology,* 1959, **59**, 146–148.

Hanfmann, E., & Kasanin, J. S. Conceptual thinking in schizophrenia. *Nervous and Mental Disease Monographs,* 1942, No. 67.

Hirschman, W. A study of the communication of learned meaningful material by schizophrenics and non-psychotics and its relationship to interpersonal involvement. *Dissertation Abstracts,* 1954, **14**, 186–187.

Kaufman, L. N. The development of a proverb scale for the measurement of thinking pathology in schizophrenia and a further investigation of the process-reactive dimension. Unpublished manuscript, University of Illinois, 1960.

Kaufman, L. N. The effects of interpersonal interaction situation upon abstract thinking performance in schizophrenia. Unpublished doctoral dissertation, University of Illinois, 1963.

Kin, D. *Dictionary of American proverbs.* New York: Philosophical Library, 1955.

Lothrop, W. W. Psychological test covariates of conceptual deficit in schizophrenia. *Journal of Consulting Psychology,* 1960, **24**, 496–499.

Lothrop, W. W. A critical review on the conceptual thinking of schizophrenics. *Journal of Nervous and Mental Disease,* 1961, **132**, 118–126.

McGaughran, L. S., & Moran, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *Journal of Abnormal and Social Psychology,* 1956, **52**, 43–50.

Payne, R. W., & Hewlett, J. H. G. Thought disorder in psychotic patients. In H. J. Eysenck (Ed.), *Experiments in personality.* Vol. 2. London: Routledge & Kegan Paul, 1960. Pp. 3–104.

Phillips, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease,* 1953, **117**, 515–525.

Rapaport, D. *Diagnostic psychological testing.* Vol. 1. Chicago: Year Book Publishers, 1945.

Salzinger, K., & Pisoni, S. Reinforcement of affect responses of schizophrenics during the clinical interview. *Journal of Abnormal and Social Psychology,* 1958, **57**, 84–90.

Schupper, F. X. Cognitive changes as a function of interpersonal conditions in two groups of schizophrenics. *Dissertation Abstracts,* 1955, **15**, 1443–1444.

Sullivan, H. S. The language of schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: University of California Press, 1946. Pp. 4–16.

Ullmann, L. P., Forsman, R. G., Kenny, J. W., McInnis, T. L., Unikel, I. P., & Zeisset, R. M. Selective reinforcement of schizophrenics' interview responses. *Behaviour Research and Therapy,* 1965, **2**, 205–212.

Ullmann, L. P., & Giovannoni, J. M. The development of a self-report measure of process-reactive continuum. *Journal of Nervous and Mental Disease,* 1964, **138**, 38–43.

Ullmann, L. P., Krasner, L., & Collins, B. J. Modification of behavior through verbal conditioning: Effects in group therapy. *Journal of Abnormal and Social Psychology,* 1961, **62**, 128–132.

Ullmann, L. P., Krasner, L., & Edinger, R. L. Verbal conditioning of common associations in long-term schizophrenic patients. *Behaviour Research and Therapy,* 1964, **2**, 15–18.

Ullmann, L. P., Krasner, L., & Ekman, P. Verbal conditioning of emotional words: Effects on behavior in group therapy. *Research Report Veterans Administration, Palo Alto,* 1961, No. 15.

Weiss, R. L., Krasner, L., & Ullmann, L. P. Responsivity of psychiatric patients to verbal conditioning: "Success" and "failure" conditions and pattern of reinforced trials. *Psychological Reports,* 1963, **12**, 423–426.

White, M. A. A study of schizophrenic language. *Journal of Abnormal and Social Psychology,* 1949, **44**, 61–74.

(Received July 16, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 363–368

SYMPTOMATIC TREATMENT OF ACQUIRED AND PERSISTENT ENURESIS[1]

JACK NOVICK[2]

New York University

2 groups of latency-age male enuretics were given symptomatic treatment consisting of routine supportive contact followed, if necessary, by the wetting-alarm treatment for enuresis. 1 group was composed of 22 persistent enuretics (those who have wet since birth) and the other of 23 acquired enuretics (those who have started to wet after a period of continence). In contrast to the persistent enuretics the acquired enuretics responded more favorably to routine contact alone; took less time; wet less frequently; and showed a faster rate of decrease in the course of symptomatic treatment. However, they had a higher rate of relapse, and had more symptoms and more deterioration in other areas of behavior in general after treatment.

Among the many types of enuretics distinguished by workers in this field, the contrast most frequently made is between persistent and acquired enuretics. Persistent enuretics are those who have wet since birth, while acquired enuretics are those who have started to wet after a period of continence. The distinction, first reported in the medical literature by Thursfield (1923), has so grown in importance as to warrant the lead editorial in a recent issue of the British Medical Journal (1960). This article represents the view of most present-day authorities and clearly states that acquired and persistent enuretics differ in etiology, personality characteristics, and prognosis. Furthermore, the authors of the article state that the nature of the treatment employed depends on whether the enuresis is of a persistent or acquired type.

Despite the importance of this distinction, there have been only four studies in which a direct comparison has been made between acquired and persistent enuretics. Lord (1952) found no significant differences between these two groups in a series of tests specifically designed for his study. Adler (1959) found that the placebo effect was significantly greater in acquired than in persistent *adult* enuretics. Hallgren (1957) compared the persistent and acquired enuretics in his sample on over 30 variables. The two groups differed on only two of the variables, with the persistent group having a higher proportion of familial cases and a higher average frequency of wetting. There is only one study which provides at least partial support for the hypothesis that the two groups differ. Gunnarson and Melin (1951) found that a significantly higher proportion of children with persistent enuresis showed abnormal EEGs than the group with acquired enuresis. In addition, the authors found that in the persistent group, heredity of enuresis was twice as frequent, and evidence of immaturity and infantile traits was found twice as often, as in the group of acquired enuretics.

There is, as yet, little evidence to support the view that the two types of enuretics are fundamentally different, and no evidence that whatever differences between them are of practical or prognostic importance. Some authorities believe that the difference between the two types is insignificant. Jones (1960) states that there is no reason to believe that the two types should respond differently to symptomatic treatment. Bakwin and Bakwin (1960), one of the foremost authorities in the field, state that enuresis, in the previously continent child, can be traced to the same etiological base as in the persistent enuretic.

---

[1] Paper presented at the annual meeting of the American Psychological Association, September 1964. This paper is based, in part, on a dissertation submitted to the faculty of the Graduate School of Arts and Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University. The author wishes to express his gratitude to his research advisor, Margaret E. Tresselt and committee members, Jacob Cohen, Bernard Kalinkowitz, Haim Ginnott, and Morris Eagle for their many helpful suggestions and encouragement during the course of this study. The enuresis training equipment was provided by the Enurtone Company of New York.

[2] Now at the Hampstead Clinic, London, England.

The present study was designed to test if children with persistent and acquired enuresis: (*a*) differ in response to symptomatic treatment; (*b*) differ in the long-term effectiveness of cure produced by this type of treatment; (*c*) differ in the after-effects of symptomatic treatment.

Symptomatic Treatment

The symptomatic treatment measures used in this study consisted of initial routine supportive treatment of all enuretics, followed, when necessary, by treatment with the wetting-alarm apparatus. Routine treatment consists of a variety of supportive measures aimed at relieving the child's guilt, increasing his confidence, and becoming his ally by dissuading the use of all punitive and harsh procedures on the part of the mother. Many authors (Anderson, 1930; Kanner, 1948; Mowrer & Mowrer, 1938) have felt that these routine measures, combined with the suggestive influence and positive attitude of the therapist, are the main factors leading to improvement. The results of research using routine supportive measures alone (Anderson, 1930) or routine measures in comparison with specific treatment (Adler, 1959; Brown & Ford-Smith, 1941) indicate that these routine measures are at least as effective in controlling enuresis as any of the specific drugs or other physical praxes employed.

The Wetting-Alarm Training Apparatus

Mowrer and Mowrer (1938) were the first to devise and, for therapeutic purposes, employ an efficient wetting-alarm apparatus for waking a child immediately following the onset of micturition. The apparatus used in this study is an improved Mowrer-type wetting-alarm apparatus devised by Sieger (1952). There are 15 studies reported in the literature on the efficacy of the Mowrer apparatus or a variant developed by Crosby (1950) which uses shock rather than a bell as a waking stimulus. The results have been uniformly positive. The research is reviewed by Jones (1960) and Lovibond (1964).

PROCEDURE

This study was conducted under the aegis of the Staten Island Mental Health Center.[3] No case was accepted without a medical examination for possible physically based bed-wetting. The study was limited to male enuretics no younger than 6 years of age, with a frequency of bed-wetting no less than 7 out of 10 nights.

Information as to whether subjects were acquired or persistent enuretics was based on repeated questioning of parents. The importance of the distinction was emphasized and parents were told that the form of treatment depended on the type of enuresis. In order to eliminate borderline cases of persistent or acquired enuresis, a detailed history of the course of enuresis was obtained and stringent criteria were employed. Persistent enuretics with a history of decreased frequency of wetting to less than once per week, maintained for 1 month or more, were eliminated from the study. Acquired enuretics were accepted for the study only when the parents could present external corroboration for a history of a minimum of 6 months completely free of any incidents of bed-wetting, and a minimum of 6 months of enuresis after onset.

After receipt of the medical referral, the study proceeded as follows:

Step I

The mother was contacted by telephone and an appointment was arranged on a date 10 days following initial contact. The distinction between acquired and persistent enuresis was explained to the mother, and the importance of accurate information was stressed. She was told to discontinue all forms of treatment and to keep daily records of the wet and dry nights. The mother was told that she would be asked about her child's behavior, fears, habits, and mannerisms, and was told to observe her child in order to increase the accuracy of her report.

Step II

Mother and child were seen at the clinic, the average interview lasting 2 hours. Mother and child were first interviewed separately. The child was measured on:

1. Degree of Anxiety as measured by the Children's Personality Questionnaire, "The CPQ" (Coan & Cattell, 1959; IPAT Handbook, 1960).
2. Intelligence as measured by Factor B of the CPQ (IPAT Handbook, 1960).
3. Extent and severity of other behavioral disorders as measured by a modification of a behavior rating scale developed by Lapouse and Monk for an epidemiological study of behavior characteristics in

---

[3] The author wishes to acknowledge his gratitude to R. M. Silberstein, medical dircetor, and W. Mandell, research director, for their support during the course of this study.

children (Lapouse & Monk, 1958, 1959).[4] The mother was questioned on: (*a*) type of enuresis, (*b*) extent and severity of other symptoms using the parents' form of the behavior-rating scale, (*c*) socioeconomic variables.

The mother and child were then brought together and routine supportive treatment was instituted.[5] Mother and child were instructed as to the method of daily record-keeping. The importance of accurate, daily recording of size and frequency of wetting was emphasized.

Step III

Ten days after the start of routine supportive treatment, the mother was contacted by telephone to determine the results of treatment and to arrange for a home visit.

Step IV

Mother, father, and child were seen in the home. If the effects of routine treatment[6] were positive, the child received further support and encouragement. If negative, the wetting-alarm training apparatus was demonstrated and installed.[7]

The method of daily record-keeping was gone over again and its importance reemphasized. A month-long record card was filled out daily by mother and child and a separate 10-day card was filled in by the mother and mailed to the clinic.

Step V

Thirty days after reaching the cure-criterion of 14 consecutive dry nights, mother and child were seen for a follow-up interview at the clinic. The child was retested on: (*a*) degree of anxiety; (*b*) extent and severity of other behavioral disorders. Further, he was administered a questionnaire related to improvement or deterioration in general areas of functioning. The mother was questioned on: (*a*) The extent and severity of other behavior disorders in her child. (*b*) Improvement or deterioration in general areas of functioning. Mother and child were given a 1-month record card and instructed to keep careful daily records of wet and dry nights.

Step VI

Every month, for a minimum of 6 months and a maximum of 1 year, the mother was contacted by telephone to check on the further progress of the child and sent a new 1-month record card.

RESULTS

Forty-five subjects were accepted for study and treatment. Twenty-three subjects were acquired enuretics with a minimum of 6 months completely dry prior to onset, and the remaining 22 subjects were persistent bed wetters who had never had a substantial or prolonged decrease in the frequency of enuresis.

There was no significant difference between the two types of enuretics on any of the descriptive variables. Thus, the two groups were similar in socioeconomic class, age, religion, income and education of parents, source of referral, intelligence, frequency of bed-wetting, and degree of maladjustment as measured by the anxiety and symptom scales. The children were mostly latency-period children of average intelligence, wetting almost every night prior to treatment. The range of ages was from 6 to 13 years, with an average age of almost 9½ years. The subjects wet approximately 9 of the 10 nights prior to treatment with a range of 4/10 to 10/10 wet nights. Seventy-two percent of the sample had wet every night in the 10-night period prior to treatment.

All of the acquired enuretics were clearly within the criteria for acceptance. The group was dry for an average of almost 2 years prior to onset, with some subjects being dry for as long as 4 years. The majority of acquired enuretics started to wet after the age of 4, some subjects starting to wet as late as 6 and 7 years of age. The average age of onset of somewhat less than 5 years of age found in the present sample is similar to the onset age of acquired enuretics found by Hallgren (1957). The acquired group had been wetting for an average of over 4 years, with 91%

---

[4] The original 45-item scale was supplemented with questions on the presence of diuresis, encopresis, reading disturbances, and episodes of fire-setting within a year of the study. The scale consists of 30 questions on a wide range of fears and worries, 7 items referring to tension phenomena such as nail-biting, sucking or chewing clothing or other objects, nose-picking, sore-picking, and chewing or sucking lips, tongue, or biting inside of mouth. The remaining items pertain to behavior such as overactivity, inactivity, restlessness, stuttering, nightmares, eating disturbance, temper tantrums, twitches, and tics. The items were designed to reveal not only absence or presence of a given behavior manifestation but also its frequency and intensity.

[5] For complete description of procedures of routine supportive treatment see Novick, 1963.

[6] There were three types of responses to routine treatment: (*a*) no effect or less than 15% decrease in frequency, (*b*) continual decrease in frequency until completely dry, and (*c*) initial decrease followed by reversion to original frequency.

[7] For complete description of procedures of wetting-alarm treatment see Novick, 1963.

TABLE 1

PERCENTAGE OF ENURETICS CURED WITH SYMPTOMATIC TREATMENT

	Persistent		Acquired		All enuretics		
	Percent cured	Number treated	Percent cured	Number treated	χ^2	Percent cured	Number treated
Routine supportive symptomatic treatment	4.5	22	34.8	23	4.67*	20.0	45
Training equipment after failure of routine treatment	85.8	21[a]	93.3	15[a]	<1.00	88.9	36[a]

Note.—Criterion for cure: 14 consecutive dry nights.
[a] Number of subjects treated by use of training equipment.
* $p < .05$.

of the sample having wet for 1 year or more after onset.

Response to Symptomatic Treatment

As indicated in Table 1, almost all the subjects responded positively to symptomatic treatment; routine supportive therapy followed if necessary by wetting-alarm treatment. Three of the 22 persistent enuretics, and 1 of the 23 acquired enuretics, treated failed to reach the cure-criterion. However, a significantly higher proportion of the acquired enuretics responded positively to routine treatment only. One of the 22 persistent enuretics, as compared to 8 of the 23 acquired enuretics, reached the cure-criterion with routine supportive treatment only.

The acquired enuretics took significantly fewer days to reach the cure-criterion than did the persistent enuretics ($t = 2.99$, $p < .01$). The persistent group required almost twice as many days of treatment (46.6) as the acquired enuretics (23.8) to reach the criterion of 14 consecutive dry nights.

Table 2 presents the results of the analysis of variance of the number of wet nights for each group prior to reaching the cure-criterion. The results indicate that the persistent enuretics had significantly more wet nights prior to reaching the cure-criterion than did the acquired enuretics. The persistent enuretics wet over twice as frequently during the course of treatment as did the acquired enuretics. The significant interaction of the analysis of variance indicates that the trend of the change in wetting in successive 10-day periods of treatment is significantly different for the enuretic groups. The trend analysis of this interaction effect indicates that the difference in trends is due to a significant difference in the linear component of the trend or the rate of decrease in wetting. The data indicate that the *rate of decrease* in wetting over successive 10-day periods of treatment is significantly greater for the acquired as compared with the persistent enuretics.

Duration of Cure or Relapse after Cure by Symptomatic Treatment

Both groups of enuretics were followed up for an equal period of time with over 80% of the total sample followed up for a period of 8 months to 1 year.

As indicated in Table 3, the acquired enuretics wet significantly more frequently after "cure" than did the persistent enuretics—on the average, more than twice as frequently. Of the 19 persistent enuretics "cured," 11 re-

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF WET NIGHTS PRIOR TO REACHING CRITERION CATEGORIZED BY ENURETIC GROUPS AND SUCCESSIVE 10-DAY PERIODS OF TREATMENT

Source	*df*	*MS*	*F*
Enuretic groups	1	56.48	11.12**
Days	6	107.88	57.38**
Groups × Days	6	4.49	2.39*
Linear component Groups × Days	2	7.54	4.01*
Quadratic component Groups × Days	2	.07	<1.00
Error I	38	5.08	
Error II	228	1.88	

Note.—Calculations based on subjects who reached criterion.
* $p < .05$.
** $p < .01$.

TABLE 3

NUMBER OF WET NIGHTS PER MONTH AFTER BEING "CURED"

Group	N	M wet per month	SD	Range wet per month	t
Persistent	19	1.98	3.3	0–12	2.02*
Acquired	22	4.83	5.6	0.2–24.0	

* $p < .05$.

mained completely dry for an average of 10 months of follow-up as compared with 5 of the 22 acquired enuretics. Of the 8 persistent enuretics with an average frequency of one or more wet nights per month after cure, 4 had an average of one or more incidents of serious wetting as compared with 14 of the 17 acquired enuretics.

After-Effects of Symptomatic Treatment

As indicated in Table 4, there is a significant difference between the two groups in the number of new symptoms, with the acquired enuretics showing more than twice as many new symptoms as the persistent enuretics. Twenty-one of the 23 acquired enuretics reported one or more new symptoms as compared with only 12 of the 22 persistent enuretics treated.

The hypothesis that the two enuretic groups would differ in deterioration in general areas of behavior is partially confirmed. The mothers of the acquired enuretics report three times as much deterioration in general areas of behavior after treatment as mothers of persistent enuretics. Seventeen of the 23 mothers of acquired enuretics reported deterioration in one or more areas of general behavior after treatment as compared with 7 of the 22 mothers of persistent enuretics. This difference is highly significant. Although the difference in deterioration as reported by children is in the hypothesized direction, it is not more than can be expected by chance (see Table 5). Relatively few of the children of either enuretic type reported deterioration after treatment.

TABLE 4

MEAN NUMBER OF NEW SYMPTOMS AFTER TREATMENT

Group	N[a]	M new symptoms	SD	Range new symptoms	t
Persistent	22	1.7	2.14	0–7	2.78*
Acquired	23	3.8	2.78	0–12	

[a] Computations based on all subjects treated.
* $p < .01$.

TABLE 5

MEAN NUMBER AREAS REPORTED AS DETERIORATED FOLLOWING TREATMENT

	Persistent ($N = 22$)	Acquired ($N = 23$)	t
Based on mother's report			
M	0.6	1.8	2.84*
SD	1.12	1.76	
Based on child's report			
M	0.2	0.4	1.25
SD	0.4	0.65	

Note.—Based on follow-up questionnaires Tapping General Areas of Behavior: relationships with peers, sibs, parents; school performance; maturity; independence; feelings of self-confidence; anxiety; and competence.
* $p < .01$.

Finally, the hypothesis that the two groups would differ in overall change in anxiety is not confirmed. Both groups showed a slight, but nonsignificant, decrease in anxiety after treatment.

SUMMARY AND CONCLUSIONS

Two types of male enuretics, 22 persistent and 23 acquired enuretics, drawn from similar homes and not differing in age, intelligence, frequency of enuresis, or degree of maladjustment were given symptomatic treatment consisting of routine supportive measures followed, if necessary, by training with the use of a wetting-alarm apparatus. In contrast to the persistent enuretics, the acquired enuretics:

1. Responded more favorably to routine supportive treatment alone.
2. Took fewer days to reach the cure-criterion with symptomatic treatment.
3. Wet less frequently prior to reaching the cure-criterion.
4. Showed a faster rate of decrease in wetting prior to reaching the cure-criterion.
5. Wet more frequently after cure based on an average 10-month follow-up.
6. Had significantly more new symptoms after treatment.

7. Had significantly more areas of general behavior reported as deteriorated by mothers.

The two groups did not differ:

1. In response to treatment with the wetting-alarm apparatus after failure of routine treatment. Almost 90% of those who failed to respond to routine treatment reached the cure-criterion when treated with the wetting-alarm training apparatus.

2. In number of areas deteriorated after treatment as reported by children. Less than a quarter of the sample of children reported deterioration after treatment.

3. In overall change in anxiety, both groups showing a slight, but nonsignificant, decrease in anxiety after treatment.

From the results of this study one can conclude that there is a fundamental difference between acquired and persistent enuretics—a difference which should be taken into account in the treatment and further study of the problem of enuresis. Specifically, the results confirm the hypothesis that acquired and persistent enuretics: (*a*) differ in response to symptomatic treatment, (*b*) differ in the long-term effectiveness of the cure produced by this type of treatment, and (*c*) differ in the behavioral after-effects of symptomatic treatment.

REFERENCES

ADLER, H. M. Enuresis in recruits: Double blind study with Equanil and review of the literature. *United States Armed Forces Medical Journal,* 1959, **10**, 767–786.

ANDERSON, F. N. The psychiatric aspects of enuresis. *American Journal of Diseases of Children,* 1930, **40**, 591–618, 818–850.

BAKWIN, H., & BAKWIN, R. M. *Clinical management of behavior disorders in children.* (2nd ed.) Philadelphia: W. B. Saunders, 1960. Pp. 363–374.

BROWN, R. C., & FORD-SMITH, A. Enuresis in adolescents. *British Medical Journal,* 1941, **2**, 803–805.

COAN, R. W., & CATTELL, R. B. The development of the early school personality questionnaire. *Journal of Experimental Education,* 1959, **28**, 143–152.

CROSBY, N. D. Essential enuresis: Successful treatment based on physiological concepts. *Medical Journal of Australia,* 1950, **2**, 533–543.

ENURESIS: Leading Article. *British Medical Journal,* 1960, **1**, 1416–1417.

GUNNARSON, S., & MELIN, K. A. The electroencephalogram in enuresis. *Acta Paediatrica,* 1951, **40**, 496–501.

HALLGREN, B. Enuresis: A clinical and genetic study. *Acta Psychiatrica et Neurologica Scandinavica,* 1957, **32**, Supplement 114.

IPAT Handbook for the IPAT Children's Personality Questionnaire "The CPQ." Champaign, Ill.: Author, 1960.

JONES, H. G. Treatment of enuresis nocturna. In H. J. Eysenck (Ed.), *Behavior therapy and the neuroses.* New York: Pergamon, 1960. Pp. 377–403.

KANNER, L. *Child psychiatry.* (2nd ed.) Springfield, Ill.: Thomas, 1948. Pp. 439–453.

LAPOUSE, R., & MONK, M. A. An epidemiological study of behavior characteristics in children. *American Journal of Public Health,* 1958, **48**, 1134–1144.

LAPOUSE, R., & MONK, M. A. Fears and worries in a representative sample of children. *American Journal of Orthopsychiatry,* 1959, **29**, 803–818.

LORD, J. P. Psychological correlates of nocturnal enuresis in the male child. Unpublished doctoral dissertation, Harvard University, 1952.

LOVIBOND, S. H. *Conditioning and enuresis.* New York: Macmillan, 1964.

MOWRER, O. H., & MOWRER, W. M. Enuresis: A method for its study and treatment. *American Journal of Orthopsychiatry,* 1938, **8**, 436–459.

NOVICK, J. Symptomatic treatment of acquired and persistent enuretics. Unpublished doctoral dissertation, New York University, 1963.

SIEGER, H. W. Treatment of essential nocturnal enuresis. *Journal of Pediatrics,* 1952, **40**, 738–749.

THURSFIELD, H. Treatment of enuresis. *Lancet,* 1923, **2**, 528.

(Received July 23, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 369–376

PERSUASION AND PRAISE AS TECHNIQUES FOR MODIFYING VERBAL BEHAVIOR IN A "REAL-LIFE" GROUP SETTING[1]

THOMAS J. D'ZURILLA

State University of New York, Stony Brook

This study investigated the effectiveness of using antecedent persuasive stimulation and response-contingent praise, within a small discussion group context, to modify the rate of verbal participation of Ss in an independent, classroom situation. Ss were 96 male college students who were divided into 12 treatment subgroups and 4 control subgroups of 6 Ss each. 3 treatment conditions were employed: (a) maximum persuasion, (b) maximum praise, and (c) neutral (i.e., minimum persuasion and praise). The treated Ss showed more criterion verbal participation than did the control Ss (.01 level). No differences were found among the 3 treatment conditions. Behavior changes in the treated Ss were attributed to the effects of nonspecific factors in the treatments.

An attempt is being made to develop an adequate psychology of (social) behavior influence and control, based on principles derived from learning theory and research, to serve as the framework for a study and reformulation of verbal psychotherapy (both individual and group) in learning terms (Krasner, 1962; Krasner & Ullmann, 1965). One relevant area of research for this purpose is the experimental modification of verbal behavior through the application of procedures that approximate the Skinnerian paradigm of operant conditioning (see reviews by Greenspoon, 1962; Krasner, 1958, 1962; Williams, 1964). In recent years an increasing number of investigators in this area have attempted to avoid the artificiality of the laboratory experiment by employing such naturalistic social settings as casual conversations (Azrin, Holtz, Ulrich, & Goldiamond, 1961; Verplank, 1955), free-responding interviews (Rogers, 1960; Waskow, 1962), and group discussions (Bachrach, Candland, & Gibson, 1961; Cieutat, 1959, 1962). Despite some thorny methodological problems (Azrin et al., 1961; Bachrach et al., 1961), the results of these studies have generally supported the findings obtained in the laboratory, that is, that one person can influence another person's verbal behavior in a specifiable direction by applying stimuli in accordance with certain learning principles—for example, reinforcement.

When considering the implications of these studies for psychotherapy, however, an important question is the degree to which the behavior changes generalize to behavior outside the treatment setting. Previous reviews of studies concerned with this issue show that the results have been unimpressive and inconsistent (Greenspoon, 1962; Williams, 1964). However, as Bandura (1961) has noted, such inconsistency is hardly surprising since the experimental manipulations in most studies have involved only one session and have been barely sufficient to demonstrate significant behavior changes in that session, let alone generalization of changes to other settings.

In most previous research in this area the emphasis has been on the use of consequent (response-contingent) stimulation which is expected to reinforce the verbal behavior under consideration, generally a verbal remark connoting praise or approval. Since the concern is primarily with the manipulation of previously learned behaviors, and not with the acquisition of novel response habits, it would seem that behavior change might be facilitated by making greater use of antecedent stimuli which would be expected to either elicit or set the occasion for the occurrence of the desired verbal behavior. One possible form of such stimulation, namely, "persuasion" (i.e., urging, prompting, suggestion, etc.) can be readily modified so as to be applicable to almost any type of previously acquired behavior. Moreover, since it is well-known as an influencing device in advertising, education, child-training, etc., its use in a treatment setting with reinforcement may also facilitate transfer of behavior changes to "real-life" situations.

---

[1] This article is based upon a thesis submitted to the University of Illinois in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology (1964). The author would like to express his appreciation to J. McV. Hunt for his direction, support, and encouragement in this research project. The author is also indebted to Jerry S. Wiggins, William M. Gilbert, Donald T. Shannon, Ledyard R. Tucker, and Merle M. Ohlsen for their suggestions and criticisms. Appreciation is expressed to Lloyd G. Humphreys and Donald E. Dulany for permitting the modifications in Psychology 100 classroom procedures which made this study possible.

The present study was designed to evaluate the effects of treatments involving both persuasion and praise on the rate of verbal participation in a classroom discussion situation. The latter (criterion) situation was independent of the treatment setting, where persuasion and praise were applied within a six-member discussion group context over a period of three sessions. Three different treatment conditions were employed. In the first, persuasion was maximized and praise was kept to a minimum. This procedure was reversed in the second condition. The third was a "neutral" condition where both persuasion and praise were reduced to a minimum. A no-treatment control group, which participated only in the criterion discussion situation, was also included in the study.

METHOD

Sixteen discussions sections of the introductory course in psychology at the University of Illinois were involved in the study. These sections contained from 19 to 25 students and met for a 50-minute period once a week throughout the semester, occupying the time from 1 P.M. to 5 P.M., Monday through Thursday. The subjects were 96 male students who were randomly selected from the 16 discussion sections. The study extended over a period of 13 weeks with the treatment and control procedures being conducted for 5 consecutive weeks beginning with the sixth week. The 5 weeks preceding treatment were used to obtain the pretreatment verbal participation measures and to administer a battery of self-report personality tests. The criterion measures of verbal participation were collected during the 3 weeks following the five weekly sessions of treatment.

Pretreatment Classroom Procedure

Seven discussion-section instructors were involved in the study. All were graduate students in psychology at the University of Illinois. No instructor was connected with more than three of the sections and only one of them had less than two sections. During the first class meeting the instructors introduced the students to the course and explained the procedures of the discussion section. All instructors were given a description of a student-centered teaching procedure (see Rogers, 1951) and asked to follow that procedure throughout the semester.

During the second, third, and fourth class periods, the experimenter was present as an observer in all discussion sections. He was introduced as "a staff member who was conducting an evaluation of course procedures." For this 3-week period, the experimenter recorded the amount of verbal participation by each student during the class meetings. Since the students were assigned permanent seats for the semester, verbal participation units were recorded on seating charts with numbered blocks representing students.

Definition of verbal participation. Included in the response class of verbal participation was any verbalization by a student which satisfied the following four criteria:

1. *The verbalization must constitute at least one complete statement.* A complete statement was defined as a sentence having a subject and a predicate and containing a complete thought or idea. An incomplete sentence was considered complete if the missing part was strongly implied.

2. *The verbalization must be emitted voluntarily.* A verbalization which was elicited by means of a direct question to a specific individual was not recorded as verbal participation. However, a voluntary response to a question directed to the class as a whole was accepted.

3. *The verbalization must be emitted by the "principal speaker."* The principal speaker was considered the person who "had the floor" or to whom the attention of the class was directed.

4. *The verbalization must be relevant to the procedures or subject matter of the course.* Verbalizations which were unrelated to the course of introductory psychology were not recorded.

The basic unit of verbal participation was simply defined as any block of verbal participation, by a single individual, which was surrounded by verbalizations of other individuals in the class. If a pause of 5 seconds or more occurred during a unit of verbal participation, two units were recorded, but only when the verbalizations on both sides of the pause satisfied the above criteria for verbal participation.

Subject Selection and Formation of Treatment Groups

Prior to the fifth week, six male students were randomly selected from each of the 16 classroom discussion sections by using a table of random numbers. Thus, 16 subgroups of six subjects each ($N = 96$) were formed. These subgroups were then assigned to four conditions with four subgroups in each condition: (*a*) Persuasion (Pn), (*b*) Praise (Pe), (*c*) Neutral (Nl), and (*d*) Control (Cl). The primary basis for assigning subgroups was the counterbalancing of conditions with respect to time (1 P.M., 2 P.M., 3 P.M., 4 P.M.) and day (Monday, Tuesday, Wednesday, Thursday).

During the fifth class period, all students in the class were asked to complete a personality test battery with the explanation that a survey was being conducted by the psychology department on the attitudes of college students toward themselves and a variety of situations and concepts.[2] At the conclusion of the fifth class period, subjects assigned to Conditions Pn, Pe, and Nl (treatment conditions) were informed by the experimenter that the psychology department was going to conduct a few small discussion groups on an experimental basis for a period of 5 weeks. Each subgroup of six subjects was told that it had been selected for such a group, which would meet with the experimenter in a separate room, on the same day, and at the same hour that its regular classroom section was meeting. No contact was made with subjects from Condition Cl since they were to remain in the classroom during the treatment period and serve as the control for the treatment procedures. With the obvious exception of Condition Cl, the instructors were unaware of the condition to which their students had been assigned; nor were they aware of the nature of the various treatments. Only one of the four conditions contained subgroups which shared a common discussion-section instructor. This occurred in Condition Pe with two subgroups.

Treatment Conditions

In each of the 12 treatment groups the six subjects were seated around a small table and asked to keep the same seats for all sessions. The procedure for the first two sessions was the same in all groups. In the first session, subjects were introduced to each other and the objectives and procedure of the group were discussed. At this time the expected group member role was described, which stressed the responsibility of each member to come prepared to the sessions with topics for discussion and to contribute his share to the discussions. In an attempt to introduce a positive attitude or "set" in regard to verbal participation, the benefits of active participation (e.g., easier learning, better grades) were emphasized. In addition, an attempt was made in the first session to create a group atmosphere which could be described as generally positive, informal, and relaxed.

In the second session, a standardized discussion procedure was begun which was subsequently followed during every session in all treatment groups. This procedure could generally be described as a student-centered approach to group discussion. Each session began with the following statement from the experimenter: "What would you like to begin with today?" Further comments were offered by the experimenter only when it was necessary to explain or clarify poorly understood material and, in the final three sessions, to implement the techniques of persuasion and praise. Moreover, the verbalizations of the experimenter were always as brief as possible and were kept to a minimum when the interaction between subjects was high. The procedures for the various treatment conditions diverged in the third session.

Persuasion (*Pn*). In this condition, the experimenter's major objective was to persuade the subjects to participate as frequently as possible in the group discussions. Persuasion was defined as a deliberate attempt by the experimenter to induce or elicit verbal participation by urging, prompting, advising, and requesting information. The following are examples of the experimenter's persuasive comments: "What do you think about that viewpoint, Tom?" "John, you should try to enter the discussion more often." "Is there anything that you would like to introduce for discussion today, Bill?" "Do you have anything to add to that, Joe?" "You will find that you actually learn more when you participate often in the discussions, Jim." "Can you tell us a little more about that experiment, Bob?" Statements such as these were used as frequently as necessary in order to maintain a high level of participation by as many subjects as possible. Every attempt was made to employ persuasion in a helpful and nonthreatening manner. In this condition, the experimenter tried to minimize all gestures and comments which might suggest praise or approval during and following the subjects' verbalizations.

Praise (*Pe*). The primary objective of the experimenter in this condition was to reward each unit of verbal participation with a comment conveying praise or approval. While the subject was speaking, the experimenter emitted such comments as "mmm-hmm," "good," and "right." Following the subjects' verbalizations, the experimenter employed statements such as "That was a good point," "That is very interesting," "Very good," "Good idea," and "That is an important question." Praise was communicated even after statements which were not entirely correct. In this case, the experimenter expressed his approval for a good or interesting try and then followed with a statement of clarification. Praise was employed by the experimenter as frequently as it was possible to do so without appearing stilted or contrived. In this condition, no deliberate attempt was made by the experimenter to induce or persuade subjects to participate.

---

[2] The battery consisted of the following instruments: (*a*) Anxiety Differential (Alexander & Husek, 1962), (*b*) a short form of the Personal Report of Confidence as a Speaker (Paul, 1966), (*c*) Pittsburgh Social Extraversion-Introversion and Emotionality Scale (Bendig, 1962), (*d*) IPAT Anxiety Scale Questionnaire (Cattell, 1957), and (*e*) selected situations from the S-R Inventory of Anxiousness (Endler, Hunt, & Rosenstein, 1962). In addition to these personality measures, total scores from the School and College Ability Test (SCAT; Educational Testing Service, 1957) were obtained and another measure of academic ability was provided by two course exams which were given on Wednesday evening during the fifth and eleventh weeks of the semester.

Neutral (*Nl*). In this condition, the procedure initiated in the second session in all treatment groups was continued for the remaining three sessions with no additional manipulations being introduced. The manner of the experimenter was positive, friendly, permissive, and interested. To all speakers, he devoted his full, positive attention but he attempted to remain as "neutral" as possible in his verbalizations throughout treatment. That is, he discussed with the subjects the subject matter of general psychology while avoiding, insofar as possible, the use of persuasion and praise in the manner described in the above conditions.

Posttreatment Classroom Procedures

At the conclusion of the five treatment sessions, all subjects returned to their regular classroom discussion sections. An assistant experimenter (who was unaware of the nature of the study) was present as an observer in all 16 discussion sections for a period of 3 consecutive weeks following the treatment period. Criterion verbal participation data were collected by the assistant experimenter in the same manner that the pretreatment measures were obtained earlier in the semester by the experimenter. In order to obtain data from which the reliability of the recording method could be determined, the experimenter recorded verbal participation units independently for four consecutive class periods.

RESULTS

Reliability of Recording Verbal Participation Units

The reliability of recording verbal participation units in the classroom was established on the basis of four discussion-section periods recorded independently by the experimenter and the assistant experimenter. The product-moment coefficient between the ratings of the two observers was .99. This figure was based on 76 students using the total number of verbal participation units for the class period as the basic response unit.

Validity of Treatment Conditions

Before turning to the treatment results it is necessary to examine the degree to which the experimenter was able to implement the types of verbal responses required by the various treatment conditions. From tape recordings of the treatment sessions, 106 units of the experimenter's verbal behavior were randomly selected. Forty-three of these units were taken from Condition Pn, 43 from Condition Pe, and 20 from Condition Nl. A unit was again defined as any block of verbal behavior which was surrounded by the verbalizations of other individuals. The 106 verbal units were then typewritten in random order and preceded by the following definitions:

Persuasion (*Pn*). The verbal unit indicates an attempt by the experimenter to induce or elicit the verbal participation of a group member or group members whether by urging, prompting, advising, or merely requesting information.

Praise (*Pe*). The verbal unit contains an overt expression that the experimenter approves of, or is pleased with, a comment made by a group member and suggests praise for a correct, significant, or interesting statement, or just for the fact that the group member had participated.

Neutral (*Nl*). The verbal unit constitutes a statement, clarification, or explanation concerning the subject matter of general psychology but contains no overt expression of persuasion or praise relative to participation.

Two judges (advanced graduate students in psychology) and the experimenter then independently classified the 106 verbal units according to the response classes Pn, Pe, and Nl. Agreement scores were obtained by determining the number of exact agreements between two individuals and dividing by the total number of units classified. The scores between the experimenter and each of the two judges were .92 and .94. The score between the two judges was .88. These results indicate that the types of verbal responses required by the three treatment conditions could be reliably differentiated.

In order to determine the degree of correspondence between the various types of verbal responses and the treatment conditions requiring those responses, a count was made of the number of correct classifications made by each judge. Since Nl responses were necessarily present in all treatment conditions, units from Conditions Pn and Pe which were classified as Nl were included in the "correct" category. The figures for the two judges—after dividing by the total number of units classified—were .97 and .96, which support the validity of the treatment conditions.

Effects of Treatment Conditions

In order to evaluate the effects of the various treatments on verbal participation in the classroom discussion situation, pretreatment and criterion scores of verbal participation were computed for each subject by summing the number of verbal participation units for class periods 2 through 4 (Pre-VP), and for periods 11 through 13 (Criterion VP). In the case of a few subjects who were absent for one of the class periods, the verbal participation score was the total number of units for the two class periods at which the subject was present plus the average number of units for those periods. A simple analysis of covariance was then performed using the mean Criterion-VP scores for Conditions Pn, Pe, Nl, and Cl adjusted according to the mean Pre-VP scores. The F associated with the adjusted mean square for Conditions was found to be significant beyond the .01 level ($F = 5.10$; $df = 3/91$). A summary of the analysis of covariance is presented in Table 1. The obtained and adjusted means are presented in Table 2.

Multiple comparisons were then made among the adjusted means by using an extension of Duncan's new multiple-range test (Kramer, 1957). This test showed that each of the treatment means (Pn, Pe, Nl) was significantly different from the control mean (Cl) beyond the .01 level, but that there were no significant differences among the treatment means.

Since a large decrease from Pre-VP to Criterion VP was found in Condition Cl (see Table 2), a 3×2 analysis of variance was performed in order to determine whether the treatment conditions significantly increased the rate of verbal participation or whether the significant F of the analysis of covariance was primarily due to the decrement in the control group. The analysis of variance contained a Treatments (Pn, Pe, Nl) main effect, a Trials (Pre-VP, Criterion VP) main effect, and a Treatments × Trials interaction. The results are summarized in Table 3.

TABLE 1

SUMMARY OF ANALYSIS OF COVARIANCE

Source	*SS*	*df*	*MS*	*F*
Conditions	491.79	3	163.93	5.10*
Error	2,924.60	91	32.14	
Total	3,416.39	94		

* $p < .01$.

TABLE 2

MEAN PRE-VP AND MEAN AND ADJUSTED MEAN CRITERION-VP SCORES

Condition	*N*	Pre-VP		Criterion VP		
				Obtained		Adjusted
		M	*SD*	*M*	*SD*	*M*
Pn	24	11.38	11.50	12.25	11.69	10.60
Pe	24	8.23	7.60	9.50	7.43	10.26
Nl	24	8.31	9.35	11.06	10.14	11.76
Cl	24	8.96	9.32	5.60	6.44	5.80

The Trials main effect was significant beyond the .05 level ($F = 4.40$; $df = 1/69$), indicating a significant increase in verbal participation for the treatment conditions. As expected, the Treatments × Trials interaction was not significant ($F < 1$).

In order to investigate whether responsiveness to the treatment conditions could be predicted from any of the personality or aptitude measures, product-moment correlations were computed between each test measure and verbal participation change scores (Criterion VP minus Pre-VP) for each of the treatment conditions. Of the 72 coefficients computed, only four were found to be significant at the .05 level, which was not greater than the number of significant correlations to be expected by chance.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE

Source	*SS*	*df*	*MS*	*F*
Treatments	222.13	2	111.06	<1
Error (a)	12,401.37	69	179.73	
Trials	84.94	1	84.94	4.40*
Treatments × Trials	34.38	2	17.19	<1
Error (b)	1,331.31	69	19.29	
Total	14,074.12	143		

* $p < .05$.

DISCUSSION

The finding that the treated subjects showed a significant increase in rate of verbal participation in the criterion situation indicates that a treatment program based on group discussion can produce a modification in verbal performance in another "real-life" group situation. Since no significant differences were found among the three treatment conditions, however, it would appear that the specific effects of persuasion and praise in this study were very slight in comparison with the effects of nonspecific factors in the treatments. The nature and source of these nonspecific factors cannot be precisely determined on the basis of the present study; however, some hypotheses based in part on previous research may be discussed.

Inspection of the experimental procedures and of a postexperimental questionnaire given to the treated subjects suggests that the most significant influencing factors were associated with the general behavior of the group leader and the nature of the group activity. The group leader singled out the experimental subjects, put them into small discussion groups, and showed much interest in their participation in the group discussions. Such special personal attention—in connection with verbal participation—was not available to the control subjects in the larger (criterion) classroom situation. With regard to the nature of the group activity, on the postexperimental questionnaire 50% of the treated subjects reported that what they liked most about the group sessions was the fact that the discussions were very interesting and stimulating. Because of the smaller size of the treatment groups, it is likely that these subjects spent considerably more time discussing topics of special interest than did the control subjects.

Within a learning framework, the factors of personal attention and interesting activity may be interpreted as having had both a (discriminative) cue function and a reinforcement function in the present study. With regard to their cue function, these factors probably set the occasion for verbal participation since it is likely that the verbal behavior of most subjects had been rewarded many times in the past in the presence of an interested and attentive "authority figure" and when the conversation or discussion was such as to be labeled "interesting." For much the same reason, these factors might have also acted as reinforcers of verbal participation. Skinner (1953) and Ferster (1961) have discussed this point in terms of the concept of "generalized reinforcement." According to Ferster, such variables as the attention of another person and simply engaging in certain activities may become powerful reinforcers without reference to a specific level of deprivation by virtue of having served as preconditions for specific reinforcements in the past.

Several experimental studies have been done which support the notion that verbal behavior may be effectively reinforced by attentional variables (e.g., looking at the subject and appearing interested, nodding, leaning forward, making expressive movements and gestures; Adams & Hoffman, 1960; Cieutat, 1959, 1962; Reece & Whitman, 1962). In addition, Hunt (1963) has reviewed many recent studies which support the view that engaging in interesting activities can act as a powerful reinforcer. Important to the interpretation of the present results is that following from this viewpoint it would be expected that such actions, once emitted, would tend to be maintained and would easily generalize to other situations, since the individual provides his own reinforcement. That the group discussions in the present study had reinforcement value is supported by the fact that many subjects perceived them to be "interesting."

Ferster points out that a generalized reinforcer "is the uniquely human reinforcer that makes possible much of verbal behavior, education in general, and self-control [Ferster, 1961, p. 267]." Extending this viewpoint to the psychotherapy situation, Krasner (1962) maintains that generalized reinforcement is the common element in all psychotherapy procedures and employs the concept to explain the so-called "placebo effects" in therapy —that is, effects due to nonspecific factors (Honigfeld, 1964). According to Krasner, most (if not all) behavior changes in psychotherapy that are attributed to "non-specific factors" are actually the result of generalized reinforcement provided by the "attention" and general manner of the therapist as he selectively responds (intentionally or unwittingly) to certain aspects of the patient's behavior. Although no definite conclusions can be reached about the nature of the influencing process in the present study, the results provide at least suggestive evidence in support of Krasner's position. On the basis of the present discussion, however, it is suggested that Krasner's analysis of the "placebo effect" should be extended to include effects due to the possible discriminative and reinforcing properties of various therapeutic activities (e.g., Rogerian, Wolpian) with respect to certain types of behavior, such as positive self-references, relaxation, etc.

From the data on reliability and validity of procedures it appears that the present method offers a useful means by which techniques for modifying verbal behavior may be applied within one group situation and their effects observed and measured in another "real-life" social situation. Some of the major methodological problems previously encountered in research in more or less naturalistic settings (Azrin et al., 1961; Bachrach et al., 1961) were avoided in the present study. It has been shown that appropriate control groups are needed to account for the effects of "nonspecific factors" and other extraneous social stimulus variables so as to avoid confounding of results. It has also been demonstrated that at least one category of verbal behavior (i.e., verbal participation as defined in this study) may be reliably identified and recorded provided that the response unit is defined in simple and precise terms and the instructions to the rater are clear and unambiguous. Finally, in the absence of mechanical recording apparatus, an independent rater who is unaware of the experimental objectives is required in order to eliminate the problem of bias in the recording of responses due to the experimenter's expectations.

REFERENCES

ADAMS, J. S., & HOFFMAN, B. The frequency of self-reference statements as a function of generalized reinforcement. *Journal of Abnormal and Social Psychology*, 1960, **60**, 384–389.

ALEXANDER, S., & HUSEK, T. R. The anxiety differential: Initial steps in the development of a measure of situational anxiety. *Educational & Psychological Measurement*, 1962, **22**, 325–348.

AZRIN, N. H., HOLTZ, W. C., ULRICH, R., & GOLDIAMOND, I. The control of the content of conversation through reinforcement. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 25–30.

BACHRACH, A. J., CANDLAND, D. K., & GIBSON, J. T. Experiments in verbal behavior. I. Group reinforcement of individual response. In I. Berg & B. Bass (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 258–285.

BANDURA, A. Psychotherapy as a learning process. *Psychological Bulletin*, 1961, **58**, 143–159.

BENDIG, A. W. Pittsburgh scale of social extraversion-introversion and emotionality. *Journal of Psychology*, 1962, **53**, 199–210.

CATTELL, R. B. *The IPAT Anxiety Scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

CIEUTAT, V. J. Surreptitious modification of verbal behavior during class discussion. *Psychological Reports*, 1959, **5**, 648.

CIEUTAT, V. J. Sex differences and reinforcement in the conditioning and extinction of conversational behavior. *Psychological Reports*, 1962, **10**, 467–474.

EDUCATIONAL TESTING SERVICE. Cooperative School and College Ability Test. Princeton, N. J.: ETS, 1957.

ENDLER, N. S., HUNT, J. McV., & ROSENSTEIN, A. J. An S-R inventory of anxiousness. *Psychological Monographs*, 1962, **76**, 1–33.

FERSTER, C. B. Positive reinforcement and behavioral deficits of autistic children. *Child Development*, 1961, **32**, 437–456.

GREENSPOON, J. Verbal conditioning and clinical psychology. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology*. New York: Basic Books, 1962. Pp. 510–553.

HONIGFELD, G. Non-specific factors in treatment. II. Review of social-psychological factors. *Diseases of the Nervous System*, 1964, **25**, 225–239.

HUNT, J. McV. Motivation inherent in information processing and action. In O. J. Harvey (Ed.), *Motivation and social interaction: Cognitive determinants*. New York: Ronald, 1963.

KRAMER, C. Y. Extension of multiple range tests to group correlated adjusted means. *Biometrics*, 1957, **13**, 13–18.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychological Bulletin*, 1958, **55**, 148–170.

KRASNER, L. The psychotherapist as a social reinforcement machine. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy*. Vol. II. Washington, D. C.: American Psychological Association, 1962.

KRASNER, L., & ULLMANN, L. P. (Eds.) *Research in behavior modification*. New York: Holt, Rinehart & Winston, 1965.

PAUL, G. L. *Insight versus desensitization in psychotherapy: An experiment in anxiety reduction*. Stanford: Stanford University Press, 1966.

REECE, M. M., & WHITMAN, R. N. Expressive movements, warmth, and verbal reinforcement. *Journal of Abnormal and Social Psychology*, 1962, **64**, 234–236.

ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *Journal of Abnormal and Social Psychology*, 1960, **60**, 247–252.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

VERPLANK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *Journal of Abnormal and Social Psychology*, 1955, **51**, 668–676.

WASKOW, I. E. Reinforcement in a therapy-like situation through selective responding to feelings or content. *Journal of Consulting Psychology*, 1962, **26**, 11–19.

WILLIAMS, J. H. Conditioning of verbalization: A review. *Psychological Bulletin*, 1964, **62**, 383–393.

(Received July 27, 1965)

ERRATUM

In the article, "An Action Test of Hypotheses Concerning the Anal Personality," by George C. Rosenwald, Gerald A. Mendelsohn, Alan Fontana, and Alexis T. Portz in the August 1966 issue, the sentence beginning on line 22 of the second column of page 306 should read: "The product-moment correlation between number of anal items solved and efficiency shift scores on the action test was $r = -.29$ ($p = .05$) indicating that PDs tend to solve more such items."

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 377–382

TYPES OF HYPNOTIC DREAMS AND THEIR RELATION TO HYPNOTIC DEPTH[1]

CHARLES T. TART

Laboratory of Hypnosis Research, Stanford University

Several types of experiences in response to suggestions to have a dream are described, namely: (a) simply thinking about something; (b) daydreaming; (c) vivid hallucinations, like watching a film; and (d) feeling "bodily located in" a "dream world." In 2 experiments, only a minority of Ss rated their experiences as dreamlike, even when hypnotized. Significant positive relationships were found between the extent to which the experiences were rated as vivid and dreamlike, and 2 measures of hypnotic depth. The variable of whether or not Ss had gone through a hypnotic-induction procedure did not discriminate among types of response. These results illustrate the danger of pseudooperational definitions of hypnosis that ignore Ss' subjective responses.

It has long been known that if a hypnotized subject is instructed to dream, many subjects will subsequently report an experience in response to this suggestion. This response is conventionally referred to as the hypnotic dream, although Barber (1960) has shown that similar responses can be obtained under conditions of waking suggestion. As has been discussed at length elsewhere (Tart, 1965), observation suggests that there are a variety of experiential responses to dream suggestions, in contrast to the unity implied by the term hypnotic dream, and that for both theoretical and practical reasons some attempt must be made to discriminate these various experiential responses. The present paper will report on two experiments in which the subjects classified their experiences into several distinct types, and in which the relationship of response type to two measures of hypnotic responsiveness or depth was investigated.

METHOD

The present data are taken from a larger study (Hilgard & Tart, 1966) of responses to suggestions following hypnotic induction versus simply asking the subjects to imagine that the suggestions were true. The large study actually consisted of two experiments, the Fall sample and the Winter sample, with slight procedural variations, noted below.

The subjects were male and female students enrolled in the introductory psychology course at Stanford who elected to fulfill part of their course requirement of participating in experiments by acting as subjects in this hypnosis study. Each subject participated in two experimental sessions, spread over 2 consecutive days.

On each subject's first day, a self-report scale of hypnotic depth or state was explained to him, as described below. He was then given one of the three experimental treatments, and then 10 suggestibility items from a slightly modified version of Form C of the Stanford Hypnotic Susceptibility Scale (SHSS; Weitzenhoffer & Hilgard, 1962).

The three treatments for the Fall sample were: (*a*) normal waking state; (*b*) instructions to imagine that the things the experimenter suggested were true; and (*c*) a formal hypnotic induction from the SHSS. On the second day all subjects were given a formal hypnotic induction prior to the SHSS items, as for group *c*'s first day.

For the Winter sample, the three treatments were: (*a*) instructions to imagine that the suggestions were true; (*b*) the same imagination instructions, but with the explicit instructions that this might produce a hypnotic state; and (*c*) a formal hypnotic induction procedure, as with the Fall sample. On the second day subjects were again assigned to one of these three conditions, rather than all being given a formal hypnotic induction.

With the Fall sample, staff members of the Laboratory of Hypnosis Research acted as experimenter-hypnotists, with no subject having the same experimenter on both days. In the Winter sample, however, the instructions, induction, and test items were tape recorded because of a finding of significant experimenter bias in the Fall sample (Troffer & Tart, 1964). In the Fall sample, experimenters were judged to have acted more "hypnotically" in the induction condition. This bias is not particularly relevant to the present material, however.

---

[1] This paper was written during the tenure of a Postdoctoral Research Fellowship from the National Institute of Mental Health, United States Public Health Service. I would also like to thank Ernest R. Hilgard for his helpful comments on this paper, and the numerous staff members of the Laboratory of Hypnosis Research who aided in collecting the data reported herein. The author is now at the Department of Psychiatry, University of Virginia.

The instructions to have a dream, Item 5 of the modified SHSS, went as follows:

> Now I am going to allow you to rest for a little while and you are going to have a dream . . . a real dream . . . just like you have when you sleep at night. When I stop talking to you very shortly, you will begin to dream. Any kind of dream may come . . . Now you are falling asleep, deeper and deeper asleep, soundly asleep. As soon as I stop talking you will begin to dream. When I speak to you again in a minute or so you will stop dreaming if you are still dreaming, and you will listen to me just as you have been doing. If you stop dreaming before I speak to you again, you will merely remain pleasantly and deeply relaxed. Now sleep and dream. Deep asleep!

The experimenter then allowed 1 minute to pass and then asked the subject to describe whatever he had dreamed about.

At the end of the second day's session for the Fall sample, each subject was asked to characterize his experience in response to the dream suggestions as follows: (*a*) just thinking about the reported topic; (*b*) daydreaming or fantasying about the reported topic; (*c*) having vivid hallucinations about the reported topic, like watching a film; (*d*) a dreamlike[2] experience, which was not only vivid, but in which the subject seemed to be located "in" a "dream world" of experience; and (*e*) some other type of experience. The subject was also asked to characterize his previous day's experience in retrospect. For the Winter sample, the subjects made this characterization of their experiences at the end of each day.

The subjects were instructed to use a simple, self-report state scale as follows:

> During the experiment we are going to do today, we are interested in the degree to which your state of mind stays the same or changes. That is, at various times, we are going to want to know what state of mind you are in. In order to make it easy and convenient for you to tell me this, I am going to teach you a rating scale for doing this. This way, when you are asked 'State?,' you'll just call out a number to indicate your state, instead of having to explain it.
>
> Here's what the numbers are to represent. *Zero* indicates that you are normally awake and alert, just as you are now. *One* indicates that you are in a sort of borderline state, that is, a condition like the one you experience as you fall asleep at night, a time when you're not quite awake and not quite asleep. There are several things that make you aware of being in a borderline state. You may find that your thoughts are drifting a lot, or perhaps that you're daydreaming, or perhaps that your mind goes blank for periods of time. Any or all of these things can occur in the borderline state. Do you know what I mean? (Experimenter reexplains if subject replies negatively.)
>
> Now a report of *two* on the scale indicates that you are in some sort of trance-like or hypnotic-like condition. You may never have experienced such a condition before, but if you get into such a state you'll recognize it as being an unusual state of mind. If this state is not only recognizable but strong, you should report *three,* and if it's very strong and impressive, report *four*. Do you get the idea? (If subject replies negatively, reexplain.)
>
> OK, let me explain how you're to do this reporting. When you are asked, 'State?,' you are to call out the first number that pops into your mind, and this will represent your state at the time. We've found that this first impression is more accurate than if you stop to think about just what the number should be. This may seem a little hard to do at first, but it will get easy as you go along. Just call out the first number that pops into your mind whenever you are asked 'State?' Of course, if you think that the first number you've called out is *way off,* you can call out a correction, but this rarely happens.

The experimenter made sure the subject understood the scale before going on. The subjects never reported four.

RESULTS

The experimenters occasionally were unable to conduct the dream inquiry with subjects due to time pressure. Usable data were obtained, however, from 58 subjects in the Fall sample and 92 subjects in the Winter sample.

Table 1 presents the distribution of the number of subjects reporting each of the various types of responses to the dream suggestions. The 2 days of each experiment are noted separately. The "no dream" column indicates that the subject simply reported that he had not dreamed when he was asked to describe what he had dreamed about. The "intermediate" column indicates the number of subjects who categorized their experience as "half-way" between two categories; these are totaled together as there were not enough of any one kind to make separate analyses feasible. For three subjects on Day 1 and two subjects on Day 2 of the Winter sample there was no classification of the reported experience on the record sheet because of the experimenter's error.

---

[2] The adjective "dreamlike" is used here because the feeling of being "bodily located in" a "dream world" of experience seemed, a priori, an essential characteristic of night dreams. To check this assumption, the author included the question, "Were you *in* the dream world or just watching the dream scene?" on dream report forms filled out by students in his psychology courses on sleep and dreaming during the 1964–1965 academic year. Of 65 dreams from male subjects and 83 dreams from female subjects, the feeling of being "in" the dream was reported for 82% and 98% of the dreams, respectively, so the assumption of this feeling being a usual (albeit not absolutely necessary) characteristic of dreams seems tenable.

There are significant differences[3] between the distributions of responses in the two experiments, but as sample differences are confounded with differing experimental procedures, these differences are uninterpretable. Ignoring treatment differences (as will be justified below), the overall means for both experiments indicate that the most frequent response to the dream suggestions was "no dream" or "thinking." Responses of "dreamlike" are infrequent, being reported only about 11% of the time.

Judging by the infrequent use of the "other" and "intermediate" categories, the subjects seemed to have little difficulty in fitting their experiences into the classification system. A detailed reading of subjects' comments made during the dream inquiry, however, indicates that the "thinking" and "no dream" categories were often confused, many subjects apparently reporting "no dream" when they had thought about something. In the remaining analyses of this paper, therefore, the "thinking" and "no dream" categories will be combined.

The "other" responses of the subjects in both samples range over a wide variety of experiences, including comparing their responses to hypnagogic imagery, pressure phosphenes, quick flashes, static images, and emphasizing the voluntary effort needed to produce an image.

Table 2 presents the distribution of responses[4] for each day of the two samples with respect to experimental treatment. Although there is a slight tendency for "dreamlike" and "hallucinatory" responses to be more frequent with hypnosis, chi-square tests of the distributions within the main categories ("dreamlike, watching a film, daydreaming, and thinking + no dream") show that the experimental treatment is not significantly related to the distribution of responses. Experimental treatment will therefore be ignored in further analyses.

If the assumption is made that the main response categories, dreamlike ("in" a dream), watching a film, daydreaming, and thinking + no dream, represent a rank ordering of response along a dimension of vividness and complexity, with "dreamlike" being the strongest response, some interesting relations emerge.

TABLE 1

DISTRIBUTION OF RESPONSES TO "HAVE A DREAM" SUGGESTION

Session	"In" a dream	Watching a film	Day-dreaming	Thinking	No dream	Other	Inter-mediate	No data	Total subjects
Fall 1963 sample									
Day 1	5	1	7	11	27	6	1	—	58
Day 2	5	5	6	8	17	14	3	—	58
Winter 1964 sample									
Day 1	12	11	26	10	19	6	5	(3)	89
Day 2	11	14	18	12	24	5	6	(2)	90
Overall means in percent, both days of both samples	11.0%	10.5%	19.5%	14.0%	29.5%	10.5%	5.0%		100%

---

[3] It is legitimate to compare the distribution of responses on Day 1 of one sample with Day 1 of the other sample, as well as Day 2 versus Day 2. If this is done, combining the "thinking" and "no dream" categories and ignoring the "no dream data" category, the distribution of responses is significantly different for Day 1 ($\chi^2 = 20.365$ with 5 df, $p < .005$) and for Day 2 ($\chi^2 = 13.705$ with 5 df, $p < .02$). Day 1 cannot be compared with Day 2 within a sample, however, because the same subjects are used on both days, and under different experimental treatments.

[4] The total Ns of Table 2 and subsequent tables do not always add up to the total number of subjects in the study. This is because occasional subjects had to be dropped from a particular analysis because they had not had a state report at the time of the dream item, had given an intermediate dream classification, etc. Also, adjacent groups are sometimes combined in the table when the N in any one cell would have been too low.

TABLE 2

TYPE OF DREAM RESPONSE VERSUS EXPERIMENTAL TREATMENT

Treatment	"In" a dream	Watching a film	Day-dreaming	Thinking + No dream	
Fall 1963 sample					
Day 1					
Hypnotic induction	2		4	7	$\chi^2 = 4.98$
Imagination	1		1	8	$df = 4$
Waking	3		2	23	$p < .30$
Winter 1964 sample					
Day 1					
Hypnotic induction	6	3	7	10	$\chi^2 = 4.70$
Imagination—Expectation of Hypnosis	5	4	9	8	$df = 6$
Imagination	1	4	10	11	$p < .50$
Day 2					
Hypnotic induction	5	9	9	16	$\chi^2 = 3.25$
Imagination—Expectation of Hypnosis	3	1	2	6	$df = 6$
Imagination	3	4	7	14	$p < .70$

The total responsiveness of a subject to hypnotic test items is conventionally used as a measure of trance depth or profundity. Ignoring experimental treatment and looking only at the total responsiveness of each subject, we find that greater hypnotic responsiveness or trance depth (sum of all items of the SHSS passed, excepting the dream item) is associated with more vivid and dreamlike responses. The mean number of items passed with respect to type of experience is presented in Table 3 for each day of both samples. Analysis of variance for the main categories indicates that there are significant differences in means at the .001 level, with the same patterning appearing on every day.

The State Reports obtained from the subjects provide an indication of how hypnotized the subjects feel, or, alternatively, another indication of hypnotic depth.[5] Table 4 presents the distribution of response types with respect to the State Report obtained at the end of the dream item, when the experimenter (or tape) said, "The dream is over. State?" Chi-square tests of the distributions of responses for each day of both experiments indicate a significant association between State Report and type of response, with higher

---

[5] Material on the relationship between objective scores on the SHSS and the self-report depth scales will be presented elsewhere (Hilgard & Tart, 1966). They are positively related, as has been found in an earlier study with a more complex self-report scale (Tart, 1963).

TABLE 3

TYPE OF DREAM RESPONSE VERSUS TOTAL SUGGESTIBILITY SCORE

	"In" a dream	Watching a film	Daydreaming	Thinking + No dream	
Fall 1963 sample					
Day 1	5.91 ($N = 6$)		4.35 ($N = 7$)	2.35 ($N = 38$)	$F = 8.36$ $df = 2,48$ $p < .001$
Day 2	7.90 ($N = 5$)	5.00 ($N = 5$)	4.25 ($N = 6$)	3.08 ($N = 24$)	$F = 7.97$ $df = 3,36$ $p < .001$
Winter 1964 sample					
Day 1	4.79 ($N = 12$)	4.09 ($N = 11$)	2.52 ($N = 26$)	1.97 ($N = 29$)	$F = 7.68$ $df = 3,74$ $p < .001$
Day 2	4.77 ($N = 11$)	4.36 ($N = 14$)	3.19 ($N = 18$)	1.78 ($N = 36$)	$F = 11.36$ $df = 3,74$ $p < .001$

TABLE 4

TYPE OF DREAM RESPONSE VERSUS STATE REPORT

State report	"In" a dream	Watching a film	Daydreaming	Thinking + No dream
Fall 1963 sample				
Day 1				
Hypnotized (3 or 2)	3		1	4 $\chi^2 = 14.72$
Borderline (1)	2		4	6 $df = 4$
Wide Awake (0)	1		2	28 $p < .01$
Day 2				
Hypnotized (3 or 2)	4	3	2	4 $\chi^2 = 11.53$
Borderline (1)	1	2	3	12 $df = 6$
Wide Awake (0)	0	0	1	9 $p < .10$
Winter 1964 sample				
Day 1				
Very Hypnotized (3)	3	1	2	1 $\chi^2 = 17.74$
Hypnotized (2)	6	4	4	3 $df = 9$
Borderline (1)	2	4	14	5 $p < .05$
Wide Awake (0)	1	2	6	10
Day 2				
Hypnotized (3 or 2)	3	5	3	3 $\chi^2 = 14.72$
Borderline (1)	6	7	10	13 $df = 6$
Wide Awake (0)	1	2	5	20 $p < .05$

State Reports being associated with more vivid and dreamlike responses. The overall significance of this relationship for the four experimental days may be assessed using the method of Jones and Fiske (1953), yielding $p < .001$. Thus subjects who report feeling more hypnotized, regardless of experimental treatment, report more vivid and dreamlike responses to the dream suggestions.[6]

DISCUSSION

The subjects of the present study were asked to make only relatively simple distinctions about the nature of their experiences in response to the dream suggestions, distinctions which will undoubtedly be refined in future work. Yet the results demonstrate clearly that such distinctions are necessary in the study of hypnotic dreams for three reasons. First, only a minority of responses are described as dreamlike by the subjects, even those who feel hypnotized to some extent, so the term hypnotic "dream" is misleading. Second, the vividness and dreamlikeness of the response is related to subjects' overall suggestibility, both waking and hypnotic. Third, this dimension of vividness and dreamlikeness is also related to the degree to which the subjects feel hypnotized. This latter finding, based on the State Reports, is not merely a reflection of subjects' responses to suggestions, as was shown in the present data (see Footnote 6), as well as elsewhere (Hilgard & Tart, 1966). Furthermore, there are a few subjects (3 out of 150) who reported a dreamlike response ("in" a "dream world") while also reporting that they felt wide awake and normal (State Report of zero), thus underscoring Barber's (1962) question about the use of the adjective "hypnotic" with respect to these dreams.

The implications for future research on the nature of hypnotic dreams, as well as on the nature of many subjective responses brought about in hypnotic studies, have been discussed elsewhere (Tart, 1965). The above results stress the most important implication, namely, that of facing the problem of differentiating the rich and complex responses that occur in the hypnotic situation. The varied and meaningful responses that occur to suggestions to dream cannot be indiscriminately lumped under the heading of "hypnotic dream." There is too large a gap today between a position of failing to discriminate among subjective-experiential correlates of responses to suggestion and elaborate psychoanalytical speculations concerning them. We very much need some good phenomenological descriptions and surveys of the phenomena to bridge this gap, and put complex theorizing on a more empirical basis.

---

[6] Although the State Report obtained just as the dream item was finished is the closest in time to the response, and presumably reflects the subject's hypnotic state during the response to the dream suggestion, it might be argued that the subject's experience in response to the suggestion contaminates his State Report, that is, he will tend to judge his state lower if he did not respond well to the suggestion. The above analysis was repeated with the State Report obtained just *before* the dream suggestion was given, and the same pattern of relationships was seen, with the combination of four tests giving an overall result of $p < .01$.

A further important finding of the present study was the lack of a significant relationship between the experimental treatment, the formally defined independent variable, and the type of response to the dream suggestions, the dependent variable. That is, whether or not a subject had been given a hypnotic-induction procedure was not predictive of his dream response. This finding, coupled with the finding that the degree to which the subject felt himself to be hypnotized was highly predictive, emphasizes a point made elsewhere (Hilgard & Tart, 1966; Tart & Hilgard, in press) about the shortcomings of a pseudooperationalism increasingly prevalent in the work of some investigators who feel that they are studying hypnosis if they have gone through a hypnotic-induction procedure. It is operational to specify whether or not the induction procedure has been gone through, but to call this hypnosis seriously distorts the traditional meaning of the term. Hypnosis is a particular state of mind as far as the subject is concerned, or intervening variable, typically associated with an increase in suggestibility, and usually brought about by a formal induction procedure. However, induction procedures often do not have any effect on the subject, and, conversely, some subjects become hypnotized, judging by their perception of their own state of mind, without an induction procedure (or the test suggestions may constitute an induction). In the present study many subjects felt hypnotized in the Waking and Imagination conditions and, conversely, many subjects in the Hypnosis condition did not feel at all hypnotized as a result of the induction. The formally defined independent variable, the hypnotic-induction procedure, interacts with the sets and characteristics the subjects bring into the laboratory in a currently unknown fashion. Thus the outcome of this interaction, the intermediate variable of hypnosis assessed by the subjects' State Reports, is the effective independent variable which is related to the subjects' responses. Studies are needed to investigate the relationship between such independent variables as induction, laboratory setting, and subject characteristics and the resultant hypnotic state, but until these relationships are understood it is inefficient and misleading to equate the giving of an induction procedure with the presence of hypnosis. The present study, then, further emphasizes the need for a more adequate phenomenology of the hypnotic state, and for some criterion other than suggestibility or misleading sorts of operationalism for the presence and intensity of the hypnotic state. The state-scale measures offer some promise of meeting such a need, and this possibility will be discussed at length in the future[7] when the data from a number of studies using self-report state scales are analyzed.

---

[7] C. T. Tart. Self-report scales of hypnotic depth, in preparation.

REFERENCES

BARBER, T. X. The necessary and sufficient conditions to hypnotic behavior. *American Journal of Clinical Hypnosis*, 1960, 3, 31–42.

BARBER, T. X. Toward a theory of "hypnotic" behavior: The "hypnotically induced dream." *Journal of Nervous and Mental Disease*, 1962, **135**, 206–221.

HILGARD, E. R., & TART, C. T. Responsiveness to suggestions following waking and imagination instructions and following induction of hypnosis. *Journal of Abnormal Psychology*, 1966, **71**, 196–208.

JONES, L. V., & FISKE, D. W. Models for testing the significance of combined results. *Psychological Bulletin*, 1953, **50**, 375–382.

TART, C. T. Hypnotic depth and basal skin resistance. *International Journal of Clinical and Experimental Hypnosis*, 1963, **11**, 81–92.

TART, C. T. The hypnotic dream: Methodological problems and a review of the literature. *Psychological Bulletin*, 1965, **63**, 87–99.

TART, C. T., & HILGARD, E. R. Responsiveness to suggestions under "hypnosis" and "waking-imagination" conditions: A methodological observation. *International Journal of Clinical and Experimental Hypnosis*, in press.

TROFFER, S. A., & TART, C. T. Experimenter bias in hypnotist performance. *Science*, 1964, **145**, 1330–1331.

WEITZENHOFFER, A. M., & HILGARD, E. R. *The Stanford Hypnotic Susceptibility Scale, Form C.* Palo Alto, Calif.: Consulting Psychologists Press, 1962.

(Received July 30, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 5, 383–388

SOME NONDEPRIVATION VARIABLES WHICH INFLUENCE THE "EFFECTS" OF EXPERIMENTAL SENSORY DEPRIVATION[1]

C. WESLEY JACKSON, JR.[2] AND JOHN C. POLLARD

School of Nursing, Western Reserve University — *Department of Psychiatry, University of Michigan*

48 university students similar to those used in experiments on sensory deprivation (SD) were surveyed to determine their knowledge about sensory deprivation, their acknowledged reasons for volunteering to be in sensory deprivation experiments, and their previous history of unusual appearing (SD-like) experiences. The responses indicated considerable knowledge of sensory deprivation, very "active" reasons for volunteering, and a variety of previous unusual appearing experiences. The results are in accord with the hypothesis that some reported effects of sensory deprivation, particularly self-reported, unusual appearing, subjective experiences, may be strongly influenced by extradeprivation variables which are part of the total experimental situation.

Any experimental study of sensory deprivation involving human beings contains certain nondeprivation or extradeprivation psychological variables which may influence some types of data. Three such variables which previously have been discussed (Jackson & Pollard, 1962) are: (*a*) the subjects' sets or expectations about what is "supposed to happen"; (*b*) the subjects' motivations to perform in accordance with, or contrary to, these expectations; (*c*) the influence of using "free-associative"-like methods of reporting effects. These variables are most relevant for subjects' self-reports of unusual appearing subjective experiences (sensory, emotional, and cognitive) during short periods of experimental deprivation, and for those other measures which appear relatively transparent to the subjects.

The effects of the first variable, experimental expectations, have been studied (Jackson & Kelly, 1962; Kandel, Myers, & Murphy, 1958; Pollard, Uhr, & Jackson, 1963; Reed, 1962) by giving subjects general information about "previously reported effects" and then exposing them to a short period of deprivation. It was assumed that giving general information about "previous subjects' experiences" made overt and explicit what is already implicit. Consequently, the first purpose of the present study was to investigate the general experimental sophistication and the specific expectations of potential deprivation subjects.

Second, since the fulfillment of expectations during an experiment is likely to be influenced by different motivations, the subjects' conscious or acknowledged reasons for volunteering for a deprivation study were elicited.

We have hypothesized that everyone probably has unusual appearing sensations, feelings, and thoughts which often are present but which, for personal and social reasons, are rarely attended to and practically never discussed under normal circumstances. However, due to the implicit and often explicit pressure to give such responses during experimental deprivation, usually unattended thoughts, feelings, and sensations are focused upon, elaborated, and reported. Therefore, the third purpose of the present study was to inquire if potential sensory deprivation subjects had had "unusual" experiences under normal circumstances.

METHOD

Forty-eight university students taking a course in psychology were given questionnaires.[3] Questions re-

---

[1] This research was supported in part by Research Grant MY-3037 and in part by Research Grant NU-00068 from the United States Public Health Service. This paper is a revision of a presentation given at the 35th annual meeting of the Midwestern Psychological Association, Chicago, Illinois, May 4, 1963.

[2] This experiment was conducted while the senior author was at the Mental Health Research Institute, University of Michigan. The authors would like to thank Lawrence Bookbinder for permission to use his students as subjects in this study.

[3] Copies of the questionnaires may be obtained from the authors.

TABLE 1

Descriptive Characteristics of 20 Male and 28 Female Subjects

Characteristic	Range	*M*	*Md*
Age	19–43	26	23
Years of college	1.5–6	4	4
Average college grades	70–95	80.8	80
Average high school grades	70–95	85	85

Note.—25 full-time and 23 part-time students; wide variety of occupations and college majors.

garding general experimental sophistication included: biographical information, previous experience as a subject, knowledge about sensory deprivation, and sources of knowledge about sensory deprivation. Specific experimental expectations were determined by requesting predictions about sensory deprivation effects. These predictions were based upon a combination of their previous knowledge and a brief description of an experimental deprivation situation.

Naturally occurring, unusual, autobiographical experiences also were solicited by additional questionnaires. Separate forms were used for each category (i.e., visual, auditory, taste, etc.). The order of these forms was randomized to control for sequential effects. In addition, half the subjects completed these autobiographical forms about unusual experiences first, while half the subjects completed them last in order to measure possible biasing effects. Care was taken to administer and collect the questions a few at a time. This was done in order to prevent, as much as was practical, any possible changing of answers to the first questions on the basis of information gained from the later questions. The entire procedure took approximately an hour and a half.

The ages, grades, and other descriptive data concerning subjects may be found in Table 1. As might be expected, there was a disproportionate number of psychology majors, 18 of the 48 subjects listing psychology as their major or co-major. Since this class was given at night there were more working adult participants than one might encounter in the usual undergraduate course in psychology. Thus, a cross section of university students was represented, similar to those who serve as subjects in experimental studies of sensory deprivation.

Results

Previous Knowledge About Sensory Deprivation

All 48 subjects denied having ever participated in a sensory deprivation (SD) study. Exactly half, however, had participated in one or more psychological experiments with a range from one to 20 experiments. Five subjects had been in one experiment, six in two, and 13 had been in three or more.

Subjects were asked, "What do you know about 'sensory deprivation,' 'perceptual deprivation,' or 'social isolation,' including what happens during a typical experiment? Put down everything you know or expect would be true about these topics." A wide variety of "SD effects" were described including: people become neurotic or psychotic, subjects can't stand it, hallucinations, increased anxiety, abnormal effects, quick fatigue, and the like. A number of subjects made general statements about the experimenter's goals or hypotheses such as: "He's testing their reactions," "looking for disability," or "testing their sensitivity." Duration and severity of deprivation were among those given as experimental variables (see Table 2). Finally, SD experiments were associated or related to space travel, concentration camps, ESP, congenital and acquired organic conditions, the Armed Forces, retarded children, mental disorder, and other provocative topics.

Next, inquiries were made about where this knowledge was obtained. Three fourths of the subjects, but not necessarily the same 36 who responded to the previous question, checked one or more sources, with textbooks and courses being the most prevalent (see Table 3).

Acknowledged Motivation to be a Subject

At this point, subjects were asked if they would volunteer to be in a deprivation experiment, being told only that it would require 12 to 15 continuous hours for which they would be paid $1.25 per hour. Twenty-eight of the 48 subjects indicated that they would volunteer (see Table 4), half of whom listed curiosity as their major motivation. Testing themselves, helping science, or receiv-

TABLE 2

Number of Subjects Who Reported One or More Items of Information about Deprivation in Each Category (36 of 48 Subjects Gave One or More Items)

Type of information	Number of subjects
Described SD conditions	19
Gave effects of SD	14
Described experimental variables or hypotheses	19
Associated SD with other topics	10

TABLE 3

NUMBER OF SUBJECTS WHO REPORTED EACH SOURCE OF INFORMATION (36 OF 48 SUBJECTS MARKED ONE OR MORE SOURCES)

Source	Number of subjects
Textbook	25
Courses	22
Newspaper	10
Journals	9
Television	8
Magazines	5
Other	4

ing money were given as other reasons. Those who did not volunteer cited a lack of time.

Fifteen of 18 psychology majors (or comajors) volunteered as opposed to 11 of 28 nonpsychology majors. Two subjects had no major. Although a greater proportion of psychology majors volunteered ($\chi^2 = 6.87$, $df = 1$, $p < .05$), there were no further differences on any of the variables between psychology and nonpsychology majors. There also were no further differences between the total group of subjects who volunteered and those who did not.

Specific Predictions About the Effects of Sensory Deprivation

The data considered thus far pertain to variables which contribute to the development of expectations prior to any participation in an actual deprivation experiment. Next, the 48 subjects were given a general description of an SD experiment and were asked to predict the various effects of deprivation. Almost every subject reported that deprivation subjects would overestimate elapsed time at 4 hours. The averages, 8.88 hours as a prediction for most subjects, and 9.34 hours for themselves as subjects, are given in Table 5 and are significantly larger than 4 hours. The two estimates were highly correlated (r equal to .82), indicating that most individuals expected that they would perform as other people did.

TABLE 4

REASONS GIVEN FOR VOLUNTEERING OR NOT VOLUNTEERING FOR A SENSORY DEPRIVATION EXPERIMENT (PARENTHESES INDICATE THE NUMBER OF ADDITIONAL SUBJECTS WHO REPORTED IT AS A SECONDARY REASON)

For volunteering		For not volunteering	
Reason	Number of subjects	Reason	Number of subjects
Curious	14 (2)	No time	16
Money	6 (1)	Fear	1 (1)
Test self	2	No interest	1 (1)
Help science	2	Lack of knowledge	1
No answer	4	No answer	1
Total	28 (3)	Total	20 (2)

By contrast, the elapsed time judgments given by 28 different subjects from a previous study (Pollard et al., 1963) who had actually experienced 4 hours of SD, indicate that these subjects were more accurate and, if anything, tended to give underestimations (see Table 6). The mean of these experimental judgments was approximately 3 hours, which was significantly different from the questionnaire predictions.

TABLE 5

AVERAGE PREDICTED ELAPSED TIMES FOR OTHERS AND FOR SELF (WHEN PREDICTING FOR HYPOTHETICAL JUDGMENTS REQUESTED AT 4 HOURS OF SD) VERSUS 4 HOURS

	N	$\bar{X}$	t	p
For others	41	8.88	5.88	$<.05$
For self	35	9.34	5.57	$<.05$

Predictions comparing postdeprivation performance to predeprivation performance on cognitive, motor, and perceptual tests were requested. Subjects predicted for others and for themselves. Approximately three fourths of the subjects indicated for each of these predictions that postdeprivation performance would be poorer than that of predeprivation. The chi-squares (12.00—others, 8.41—self, $df = 1$, $p < .05$) were significant. There was no difference between the predictions for others and for themselves (37 same, 11 left blank).

Perhaps the most interesting and important predictions concerned what they thought would happen during deprivation. There was no apparent difference between predictions for others and for themselves, and consequently responses to both questions, except for duplications, were combined. Forty-five subjects

TABLE 6

Average Predicted Elapsed Times for Others and for Self (When Predicting for Hypothetical Judgments Requested at 4 Hours of SD) versus Actual Judgments of SD Subjects at 4 Hours of SD

	Questionnaire predictions	Actual SD judgments	t	p
	$\bar{X}$	$\bar{X}$		
For others ($N = 41$)	8.88	2.95 ($N = 28$)	5.84	$<.05$
For self ($N = 35$)	9.34	2.95 ($N = 28$)	5.87	$<.05$

predicted one or more effects including: delusions, disorganized thoughts, fears, anxieties, hallucinations, seeing beautiful dancing girls, hearing melodies, body seeming to be of abnormal size, tasting food, smelling liked or disliked things (see Table 7).

It should be noted that of the 12 subjects who reported no sources of information about SD prior to receiving a description of an SD experiment during the present study, four predicted one or more visual effects, four auditory, one taste, three touch, five emotion, and 10 cognitive. None predicted olfactory effects. Eleven of the 12 subjects predicted one or more effects. Both the patterning, with respect to selecting categories of effects, and the individual contents were quite similar to those for the subjects who reported previous sources of information about SD. For example, if the seven categories are ranked first according to the number of subjects who used each category from the group who reported no previous information about SD, and then ranked again for the 36 subjects who did report such information, the resulting Spearman rank correlation is .93, $p < .05$.

TABLE 7

Number of Subjects Who Predicted One or More Effects of SD in Each Category (45 of 48 Subjects Predicted Effects in One or More Categories)

Category	Number of subjects
Cognitive	35
Feeling	23
Visual	18
Auditory	17
Touch—Bodily	12
Taste	7
Smell	9
Complaints	5
General Sensory	9
Actions	12

The final data to be considered are the "unusual" autobiographical experiences. Half of the subjects filled out these questionnaires near the beginning of the experiment and half at the end. There were no significant order effects either on the SD predictions or on the reported autobiographical experiences.

As is indicated in Table 8, experiences were reported in all seven categories with the largest number in the cognitive or thinking category. By far the most prevalent cognitive reports had to do with ESP, with predictions of things to come, and with thoughts that something was being reexperienced when the subjects knew that this was not possible (*déjà vu*). For example:

TABLE 8

Number of Subjects Who Reported One or More Unusual Experiences in Each Category (38 of 48 Subjects Reported One or More Experiences)

Category	Number of subjects
Visual	10
Auditory	16
Taste	5
Smell	8
Touch	6
Feeling	12
Cognitive	26

> Our family has a tradition of ESP experiences which we almost take for granted and act upon as readily as we do upon sensations of the five other senses. About 2 years ago my mother had an auto accident while visiting my brother. I felt she had been seriously hurt in strange surroundings and called my sister. She told me that my mother had gone to visit my brother, and I was able to trace their actions until I found out about the accident. My wife accused me of insanity at the time—"all of those long distance phone calls."

From another subject, "Like numerous others, I periodically have the feeling that I've been here before even though it is quite impossible." Other thoughts were reports of an impending accident, a prediction of death, or the idea that one was a robot controlled by an omnipotent evil force. A particularly complex cognitive and sensory experience was described as follows:

> I was waiting for a close friend to come from the dentist's office for his usual appointment. Concentrating on this experience and thinking deeply about him I suddenly seemed to be in a strange dental office myself. I smelled the office. I saw the walls. I felt a deep twinge in my own jaw—and knew that he had had a tooth pulled. Noting the time I later checked with him. He had been unexpectedly having a tooth pulled at the very time I felt it in my own jaw and the position of the tooth was the same. The walls (color) and odor were those of his dentist's office and different from any I had ever seen before.

Feelings which the subjects were unable to explain consisted mostly of irrational fears, guilt, mood swings, depression, and anxiety. Visual reports included seeing people where there were only shadows, seeing the ceiling go round and round, or the face of an animal or person bearing the characteristics of a deceased loved one. A very striking experience was reported as "I once had a professor who literally seemed to glow white light when he would get particularly engrossed in a lecture. I saw this about three times on only this one person."

Auditory experiences were described as hearing apparently nonexistent music, footsteps, air raid sirens, clocks ticking, and even the whirling noise of the atoms and molecules of the air. Examples from other categories were: the feeling that bugs or ants were crawling over one, feeling one's body as if it were some other person, experiencing tastes and smells without a known stimulus, and the like.

DISCUSSION

University students drawn from the present sample would begin sensory deprivation, having been subjects in other types of experiments, and with specific expectations. Initially, they were able to describe SD conditions, to give effects of SD, and to list experimental variables or hypotheses. Following a brief description of experimental deprivation, they predicted overestimations of elapsed time judgments, a wide variety of unusual appearing experiences, and poorer postdeprivation test performance. In addition, their stated motives for volunteering reflect how "actively" they probably would be participating, and are the same reasons as those given by actual SD subjects (Wexler, Mendelson, Leiderman, & Solomon, 1958).

The "unusual" experiences reported are in accord with the hypothesis that selectively unattended sensations and thoughts may be elicited during deprivation, facilitated by different expectations, a permissive atmosphere, and a free-associative-like method of reporting. Such autobiographical experiences are quite common as substantiated by Galton (1883) and McKellar (1957) among others, and McKellar has pointed out their similarity to experiences reported during experimental SD. Also, Freedman and Greenblatt (1960) found a significant relationship between a history of hypnagogic imagery (one of the most common types of "unusual" experiences) and reports of imagery in experimental SD.

The sophistication of university subjects is important for experiments, such as the one by Zuckerman and Cohen (1964), which purportedly fail to show the influence of sets, expectations, and other forms of implicit suggestion during experimental sensory deprivation. When Zuckerman and Cohen compared three groups of SD subjects given varying amounts of explicit information about SD, with a supposedly naive control group, the results generally were not significantly different as all groups gave approximately the same number of effects. The point, of course, is that the "control" subjects were also sophisticated, a position which is amply supported by reviewing the instructions which these subjects received and the procedures which they experienced.

"Suggestion," as it is used in this context, is very similar to what Orne (1962) calls "demand characteristics" of experiments. While such variables are important for research on sensory deprivation, it is not contended that there is a linear relationship between amount of information about SD and experimental effects, or even that there is necessarily a one-to-one relationship between specific expectations and experimental behavior. In the present experiment, the predictions of subjects who reported no previous knowledge of SD were quite similar to those with such knowledge after both had been given a brief description of an SD experiment. In addition, predicted elapsed time estimations were overestimated as compared to the underestimations of actual SD subjects. During post-SD interviews most of the 28 actual SD subjects (Pollard et al., 1963) said that they felt as if they had been in there for a very long time, but believed that they were exaggerating and consequently tried to adjust their time estimations accordingly.

"Suggestion," or "demand characteristics," probably consists of numerous variables, and, if the literature on different measures of suggestibility is any indication (Stukat, 1958), such variables may not be highly correlated. Actually very little is known about the types of loading or suggestion involved in the "demand characteristics" of experiments, how much of any given type is enough to influence the data, and how the influence of types and amounts varies with different measures and different people. In this context, it should be remembered that many SD effects, and especially self-reported sensory experiences, are very inconsistent from subject to subject and from experiment to experiment (Jackson & Pollard, 1962). Perhaps only one third of the subjects in a given experiment would need to be influenced in the same direction in order to produce results which would be quite comparable to studies purportedly demonstrating the effects of SD.

What can an experimenter do in this situation? It is extremely difficult, if not impossible, to conduct an SD experiment without providing suggestive information. Very likely the problem is most acute when the data consist of self-reports of subjective experiences or scores from measures whose purpose is relatively obvious. One can partially cope with the situation by systematically regulating information the experimenter intentionally gives to the subjects, by, as Freedman and Greenblatt (1960) suggest, using more deceptive measures, and by including a suggestion, nonsensory deprivation comparison group as illustrated by Orne and Scheibe (1964).

The overall significance of the present study and of related studies is not to "explain away" the results of experiments on sensory deprivation. Rather, it is important to differentiate better between those effects which are clearly a function of SD and those which only appear to be, or which may be caused either by SD per se or by nondeprivation variables which are an intrinsic part of the experimental situation.

REFERENCES

FREEDMAN, S. J., & GREENBLATT, M. Studies in human isolation II. Hallucinations and other cognitive findings. *United States Armed Forces Medical Journal,* 1960, **11**, 1479–1497.

GALTON, F. *Inquiries into human faculty*. New York: Macmillan, 1883.

JACKSON, C. W., JR., & KELLY, E. L. Influence of suggestion and subjects' prior knowledge in research on sensory deprivation. *Science,* 1962, **132**, 211–212.

JACKSON, C. W., JR., & POLLARD, J. C. Sensory deprivation and suggestion: A theoretical approach. *Behavioral Science,* 1962, **7**, 332–342.

KANDEL, E. J., MYERS, T. I., & MURPHY, D. B. Influence of prior verbalization and instructions on visual sensations reported under conditions of reduced sensory input. *American Psychologist,* 1958, **13**, 334. (Abstract)

MCKELLAR, P. *Imagination and thinking*. New York: Basic Books, 1957.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

ORNE, M. T., & SCHEIBE, K. E. The contribution of nondeprivation factors in the production of sensory deprivation effects: The psychology of the "panic button." *Journal of Abnormal and Social Psychology,* 1964, **68**, 3–12.

POLLARD, J. C., UHR, L., & JACKSON, C. W., JR. Studies in sensory deprivation. *Archives of General Psychiatry,* 1963, **8**, 435–454.

REED, G. F. Preparatory set as a factor in the production of sensory deprivation phenomena. *Proceedings of the Royal Society of Medicine,* 1962, **55**, 1010–1014.

STUKAT, K. G. *Suggestibility, a factorial and experimental analysis.* Stockholm: Almquist & Wiksell, 1958.

WEXLER, D., MENDELSON, J., LEIDERMAN, P. H., & SOLOMON, P. Sensory deprivation: A technique for studying psychiatric aspects of stress. *American Medical Association Archives of Neurology & Psychiatry,* 1958, **79**, 225–233.

ZUCKERMAN, M., & COHEN, N. Is suggestion the source of reported visual sensations in perceptual isolation? *Journal of Abnormal and Social Psychology,* 1964, **68**, 655–660.

(Received September 24, 1965)

Four 1966 Psychology Texts From Prentice-Hall

THE ABNORMAL PERSONALITY THROUGH LITERATURE

Edited by Alan A. Stone, McLean Hospital, Massachusetts, and Sue S. Stone. " . . . (this book) will attempt to present the 'outward forms' of abnormal personality utilizing the portraits drawn in literature. None of these portraits are as stereotyped as the ordinary case histories presented to illustrate abnormal personality types. They are, however, closer to the reality of human nature, which resists oversimplified classification. These examples from fiction hopefully will permit the reader to grasp the traditional classification of psychological disorders as they can be identified within vivid delineation of human lives and human relationship."—*from the Introduction*

January 1966, 423 pp., $4.50

PSYCHOPATHOLOGY OF CHILDHOOD

by Jane W. Kessler, Director, Mental Development Center, Western Reserve University. ". . . obviously written from an attitude of personal experience in clinical practice and teaching, it is richly laced with case study materials. It is not restricted in scope to behavior problems and emotional disturbances, but takes in a somato-psychological, psychosomatic, language, educational and intellectual deficits and difficulties as well."—*from a pre-publication review*

February 1966, 544 pp., $8.25

PERSONALITY RESEARCH: A BOOK OF READINGS

Edited by Donn Byrne, University of Texas, and Marshall L. Hamilton, Washington State University. Here is a representative picture of current research, theoretical formulations and issues typical of the field of personality—with emphasis on the field as a science. This collection of readings can be successfully used independently or with basic personality texts; it is most ideally suited for use in conjunction with *An Introduction to Personality: A Research Approach,* described below. Includes: articles by Cronbach, Auld, Meehl, Sears, Harlow, MacKinnon, Mednick, Byrne, Taylor, Rogers, and Holt.

March 1966, 411 pp., paperbound, $5.50

AN INTRODUCTION TO PERSONALITY: A RESEARCH APPROACH

by Donn Byrne, University of Texas. An Introduction to Personality facilitates an understanding of the way in which the scientific method is utilized in studying behavior. Concentrating on current aspects of research which deal with personality as a science rather than a type of speculative philosophy, this new text discusses where ideas originate, how these ideas are translated into research, and how theoretical speculations and research findings are interconnected.

April 1966, 552 pp., $7.75

(PRICES SHOWN ARE FOR STUDENT USE.)

For approval copies, write: Box 903

Prentice-Hall, Englewood Cliffs, N. J. 07632

The American Psychological
Association announces . . .

a "Special Issue" of the

AMERICAN PSYCHOLOGIST

November 1966

The November issue of the *American Psychologist* will be a SPECIAL ISSUE entirely devoted to a consideration of the problems involved in the exchange of scientific and technological information within the "scientific community." The contributors are individuals who have distinguished themselves in such areas as psychology, sociology, physics and library science and who were selected so that the issue would present a variety of viewpoints in addition to those of the behavioral and social scientist. Among them are included Donald deB. Beaver, William D. Garvey, Belver C. Griffith, Don Marquis, Herbert Menzel, William J. Paisley, Edwin B. Parker, Derek J. de Solla Price, Philip J. Siegmann and Don R. Swanson.

Briefly, the articles will deal with topics relative to a) the use of data from the behavioral sciences in the improvement of techniques for the exchange of scientific information b) systems of information documentation, retrieval and transfer, and the impact and potential of automation in structuring of such systems and c) the relationship which should exist among science, technology and enterprise in order to derive maximum social utility from science.

The November SPECIAL ISSUE will be overprinted so that additional copies may be ordered from:

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

Oxford University Press

Perspectives in Psychopathology:

Readings in Abnormal Psychology

Edited by **JAMES O. PALMER** *and* **MICHAEL J. GOLDSTEIN,** University of California, Los Angeles

"Exceptionally well-organized collection of highly significant studies and approaches to the understanding of psychopathology. I am also using *The Experience of Anxiety*."—Lars P. Peterson, Minnesota State College, St. Cloud

1966 448 pp. 44 illus. paper $4.00

Principles of Group Treatment

By **ERIC BERNE,** M.D., McAuley Clinic, San Francisco

An attempt is made to mold group treatment into a more rigorous and less empirical discipline, in this first systematic treatise on the use of transactional analysis in groups. "The most lucid and useful single book in the field."—Roland G. Tharp, University of Arizona

1966 384 pp. $6.50

An Introduction to Psychopathology

Second Edition

By **D. RUSSELL DAVIS,** Bristol Royal Infirmary

Of value to all who are professionally concerned with problems of mental disorder in children and adults, this new edition serves as a bridge between the psychological laboratory and the psychiatric clinic, describing both research into clinical problems and practical applications of laboratory work.

1966 164 pp. paper $3.75

Foundations of Psychopathology

By **JOHN C. NEMIAH,** M.D., Harvard Medical School and Massachusetts General Hospital

"A substantial and well-balanced review, with an imaginative and original integrating influence upon the whole."—Louis Jolyon West, University of Oklahoma School of Medicine

1961 352 pp. cloth $6.50 paper $2.25

The Experience of Anxiety: *A Casebook*

By **MICHAEL J. GOLDSTEIN** *and* **JAMES O. PALMER**

"Excellent case histories, well-written and organized, with questions well-designed to teach the student the process of drawing conclusions from clinical material."—Clifford H. Swensen, Purdue University

1963 304 pp. paper $2.25

Oxford University Press / 200 Madison Ave. / New York, N.Y. 10016

A Unique Theoretical Approach to the Causes and Treatment of Crime, Delinquency and Deviation

CRIME AND SOCIAL DEVIATION

by Shlomo Shoham

"An important and perceptive contribution to criminology, social psychology and philosophy."

—*from the Preface by Hermann Mannheim*

Available from your bookstore or directly from the publisher at $12.50 a copy.

HENRY REGNERY COMPANY
114 W. Illinois St., Chicago, Ill. 60610

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. Scientists writing for publication will find the Publication Manual an indispensable guide.

Price $1.00

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N.W., Washington, D.C. 20036

Volume 71 December 1966 Number 6

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute and Columbia University

KURT SALZINGER
New York State Psychiatric Institute and Brooklyn Polytechnic Institute

JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR	MARY C. MCGARRAGHY	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1966 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 389–399

CAMPUS TRAGEDY:

A STUDY OF STUDENT SUICIDE[1]

RICHARD H. SEIDEN

School of Public Health, University of California, Berkeley

Prior studies of college suicides have neglected the need for an adequate comparison or control group. To remedy this situation, student suicides were compared to their nonsuicidal classmates on selected demographic variables. Suiciding students could be significantly differentiated from their fellow students on the basis of age, class standing, major subject, nationality, emotional condition, and academic achievement. The suicidal students presented similar prodromal patterns which were precipitated by scholastic anxieties, concern over physical health, and difficult interpersonal relationships. Contrary to general belief, the greatest suicidal activity occurred during the beginning, not the final, weeks of the semester. On the basis of changes transpiring in the college population, a future increase of student suicide was predicted.

The act of self-destruction rudely challenges our supposed love for life and fear of death. It is always a puzzlement, but in no case is suicide more shocking or bewildering than it is in the college student. For here are a relatively privileged group of persons enjoying valued advantages of youth, intelligence, and educational opportunity. Why should persons, seemingly so rewarded, seek to kill themselves, and, indeed, to commit suicide at a rate significantly in excess of their noncollege peers (Bruyn & Seiden, 1965, p. 76)?

This perplexing question—"Why do students suicide?"—has motivated a great deal of concern among college health authorities leading to several studies and evaluations of the problem in American universities (Braaten & Darling, 1962; Jensen, 1955; Parrish, 1957; Raphael, Power, & Berridge, 1937; Temby, 1961). Unfortunately, these studies have all had an exclusively descriptive approach. They have drawn conclusions about certain characteristics of suicidal students but, seemingly, without appreciation for the degree to which these same characteristics are shared by the entire student body population. What has been conspicuously omitted is a baseline—a standard of comparison against which the diagnostic value of their findings might be judged. One is reminded of the gentleman who, when asked, "How is your wife?" astutely responded, "Compared to what?" This very question of relative comparison must also be asked in the study of student suicides.

The present study attempted to remedy this situation by applying a reasonable standard of comparison, namely, the great majority of fellow college students who do not commit suicide. By investigating what characteristics significantly differentiate suicidal students from their classmates plus examining those situational-temporal conditions associated with campus suicides, it was hoped to achieve a clearer diagnostic picture. Once the high-risk, suicide-prone student can be identified, a large and necessary step will have been taken toward the ultimate objective of effective prophylaxis.

METHOD

The approach used in the present study was one of analytic epidemiology, that is, comparing for particular characteristics the subset of student suicides with the total student body population from which they were drawn. This particular procedure meets the methodological criteria for selection of comparison groups, as stated by MacMahon, Pugh, and Ipsen (1960):

---

[1] Revision of a paper presented to Psi Chi colloquium, Western Psychological Association, Honolulu, June 1965.

This research was supported by Grant #5 T1 MH-8104 from the National Institute of Mental Health.

> A comparison group is a group of unaffected individuals believed to reflect the characteristics of the population from which the affected group was drawn. Ideally the comparison group should not differ from the affected group in any respect (other than not being affected) which might be likely to influence the frequency of the variable or variables suspected of being causally connected. This means either that both the patient and comparison groups must be representative of the same population or that if selective factors enter into the choice of the patterns, the same factors ought to enter into the selection of the comparison group [p. 235].

The method of the present study involved a comparison of the sample of 23 University of California at Berkeley (UCB) students who committed suicide during the 10-year period 1952 through 1961, with the entire UCB student body population during this same decade. The objective of this comparison was to determine what special characteristics differentiated the suicide-prone student from his classmates. Within this framework the following working definitions were employed: (*a*) *Student*—the definition of a student was established by registration on the Berkeley campus of the University of California, in either graduate or undergraduate status, during the regular college semester periods. Summer sessions were not included because of the unreliability of data for these periods and changes in the usual composition of the student body population during summer sessions. (*b*) *Suicide*—refers to a completed suicide, established by a death certificate stating suicide as the legal cause of death. In one instance, involving a jump from the Golden Gate bridge, this was not possible. Since the body was never recovered, a certificate was not issued; however, the case was well-documented in police and newspaper files. By keeping to this legalistic definition of suicide, one runs the very likely probability that the true number of suicides will be underenumerated. For example, cases of equivocal student deaths, such as by falls or drowning, were regarded as accidental, in keeping with the coroner's findings, even though these deaths, listed as accidents, could have been suicides which were covered up to avoid the social stigma related to suicide. Indeed, it has been estimated that only about 70% of successful suicides are ever recorded as such (Dublin, 1963, p. 3). The advantage in using this definition is that one can be quite certain that deaths recorded as suicide are bona-fide cases since the error is, almost always, in the direction of underreporting. (*c*) *Exposure to risk*—the period of exposure to risk comprised the 10-year span 1952–1961 inclusive, a total of 10 academic or 7½ calendar years. This important variable, the length of exposure, was to some degree controlled since both the suicidal and nonsuicidal students were exposed to the same period of risk. (*d*) *Population at risk*—population at risk was the total student body of UCB during the 10-year period cited. Case finding procedures were extremely painstaking, requiring several months of effort to detect and verify 23 bona-fide study cases. Numerous sources of information were used, but for the suicidal students the primary source was the standard death certificate, obtained from the state health department. Secondary sources consisted of newspaper clippings, police files, and University records. The source of materials for the baseline data for the total student body population was the UCB Office of the Registrar. Their publication, *A Ten-Year Survey of Certain Demographic Characteristics of the Student Population* (Suslow, 1963), was indispensable.

In terms of research design, the procedures consisted of collecting and analyzing data regarding selected attributes of the total student population. These data were then used as a baseline to which the sample of suicidal UCB students could be compared. Since suicide may also involve a strong volitional component, further analyses were made with respect to certain situational-temporal features of the academic environment.

RESULTS AND DISCUSSION

Results are presented in tabular and graphic form and discussed in the text by order of their appearance. The various comparisons were statistically analyzed by testing the significance of the difference between two proportions (Hill, 1961, pp. 122–132), specifically, the significance of proportional differences between the suicidal sample and expected population values as based upon knowledge of the student universe. All probability statements are two-tailed probabilities.

Incidence and Prevalence

Previous research on the UCB population (Bruyn & Seiden, 1965) investigated the general question of student suicide risk. By comparing the student suicide experience with the suicide incidence among a comparable group of non-college-age cohorts, it was established that the incidence of suicide among students was significantly greater than for non-student-age peers ($p = .004$). Conversely, the general mortality experience from all causes was significantly more favorable for students when compared to their non-academic-age peers ($p < .001$). In terms of total mortality, suicides accounted for 23 of the 68 student deaths which occurred during the 10-year study period. Proportionally, it ranked as the second leading cause of death (34%), exceeded only by accidents (37%).

TABLE 1

SELECTED DEMOGRAPHIC CHARACTERISTICS OF SUICIDAL AND NONSUICIDAL STUDENTS, UCB, 1952–61

Demographic characteristics	Suicidal students		Total student body population	p
	Frequency distribution ($n = 23$)	% distribution	% distribution	
Age				
Under 25	9	39	70	.001
25 and above	14	61	30	
Class standing				
Undergraduate	12	52	72	.033
Graduate	11	48	28	
Sex				
Male	17	74	67	*ns*
Female	6	26	33	
Marital status[a]				
Married	3	14	23	*ns*
Never married	19	86	77	
Race				
White	20	87	89	*ns*
Nonwhite	3	13	11	
Religion				
Protestant	15	65	60	*ns*
Jewish	5	22	18	
Catholic	3	13	22	
Nationality				
U.S.A.	19	83	96	.002
Foreign	4	17	04	
Major subject[b]				
Mechanical-mathematic	10	50	64	*ns*
Aesthetic-social	10	50	36	
Grade-point average[c]				
Above average	14	67	50	*ns*
Below average	7	33	50	
Mental health service				
Psychiatric patient	8	34	10	<.001
Nonpatient	15	66	90	

[a] Excludes one divorced student.
[b] Excludes three students who had not declared majors.
[c] Excludes two students who did not complete a semester.

Age

For the United States as a whole, there is a well-documented positive correlation between age and suicide (Dublin, 1963, p. 22). This same relationship holds for the student population. If the student body is divided on the basis of those who are above and below age 25, one finds that the percentage of suicides in the older age group is approximately twice their proportional percentage in the population (see Table 1). This distinction is graphically portrayed in Figure 1 which presents the relative frequency of suicidal and nonsuicidal students by 5-year age groups. It is notable that only about 6% of all students fall in the 30 to 34-year age category while more than 26% of the suicidal students are found in this interval. In fact, the median age for the student body population is 22 years, 6 months, while the median age for the suicidal students, 26 years, 5 months, is greater by almost 4 years.

Class Standing

Directly correlated with, and, indeed, almost identical to, chronological age, is the class standing of individual students. Median class standing for the entire student population was the junior year, for the suicidal subset it was the senior year. When the groups are divided on the basis of graduate or undergraduate standing, one finds that graduate students committed suicide in numbers significantly greater than could be expected from their proportions in the student body at large (see Table 1).

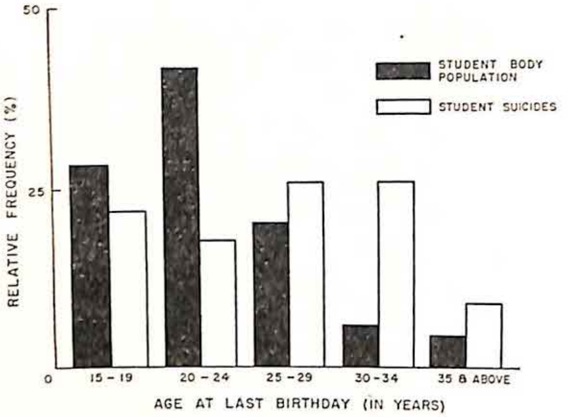


FIG. 1. Age distributions of student suicides and total student body population, UCB, 1952-61.

Sex

Of the 23 student suicides, 17 were male, 6 female, a sex ratio approximating 3:1 (see Table 1). This finding accords with those sex ratios reported in previous studies of completed suicide (Dublin, 1963, p. 23). However, an adjustment is necessary to correctly relate this information to the college population. Whereas the sexes are about equally distributed in the general United States population, they are not equally distributed on campus. For the years under study, males outnumbered females in the student body population by approximately 2:1. Accordingly, the obtained sex ratio of 3:1 must be halved to yield an adjusted student ratio of about 1.5 male suicides for each female suicide. This student sex ratio is considerably narrower than the sex ratio for the country at large. It seems to indicate a heightened risk of suicide among female students as compared to the general female population. However, this indication must remain somewhat speculative since the female suicides were considerably older (median age 30 years, 1 month) than were male suicides (median age 26 years, 1 month). As a consequence one cannot be entirely sure that the constricted ratio is not an effect of confounding between age and sex. Should further research confirm that there is, in fact, a greater risk of suicide among female students as opposed to female nonstudents, it would follow the predictions of Gibbs and Martin (1964). They proposed a rise in female suicides due to increasing social pressures. According to their status-integration theory, as more women enter the labor force they encounter cross-pressures from conflicting social roles. They postulate that these stresses will lead to increasing numbers of female suicides.

Marital Status

Of the 23 student suicides, it was possible to classify 22 persons into the categories of "married" or "never married," which corresponded to the available student population data. One divorced student was thereby excluded from the analysis. There was no remarkable disparity between the suicidal and nonsuicidal students on the basis of marital status (see Table 1). For the entire United States population, suicide is less common among married persons (Dublin, 1963, p. 26), but this was not the case for campus suicides. Only three of the student suicides were married, and only one of those married had children. The remaining two cases, both females, committed suicide shortly after their marriages.

Race

Of the 23 known suicides, only three were nonwhite and all three of these nonwhite students were Chinese. There were no suicides among Negro, East Indian, or American Indian students who, at any event, comprised only about 3% of the student body population. The distribution of suicides by race corresponded closely to the racial proportions found in the student population (see Table 1). It should be mentioned, however, that there is good reason to question the adequacy of these racial data. Since University records do not ask for nor indicate students' race, these breakdowns, furnished by the University Dean of Students Office, were presumably obtained from simple headcounts with all the imprecision that this method implies.

Religion

Religion was not a significant factor in differentiating suicidal students from the general campus population (see Table 1). As was the case with racial statistics, the religious data, likewise, must be regarded with great skepticism. The University does not conduct a religious census of its students. Consequently, the religious population figures were estimated from student residence information cards on which "religious affiliation" is an optional item. Very frequently it is left unanswered.

Nationality

Only 4 of the 23 student suicides were foreign students. Nonetheless, their representation in the student body was so negligible (only 4%) that they appear among the suicides in approximately four times the magnitude one would expect from their proportions in the student population (see Table 1). As a group, these four "international student" suicides were characterized by some striking similarities. As youngsters, all of the four had known and suffered from the ravages of war, and three of them were forced to flee from their childhood homes. Two of the students, natives of mainland China, had been dispossessed by the Communist revolution; another student, born in Austria, lost his family in the horrors of the Nazi concentration camps and subsequently migrated to Israel. The fourth student, a native Israeli, had grown up amidst the Arab-Jewish war over the Palestine partition.

Moreover, they shared a similar pattern of conflicts, centering to a large degree around strong feelings of shame. These feelings were reflected in a deep dread that they would not meet expectations that others had set for them. There was some reality to these fears, in that other persons had sent them abroad, were paying their expenses, and probably did expect from them some measure of academic achievement. Still, their excessive concern about "what others would think" was unduly frenetic. All four of them were known to the Student Mental Health Service where they had been seen for psychiatric treatment. These findings, however, must be interpreted with some caution since the median age of foreign students (26 years, 1 month), exceeded the median age of American students (24 years), raising the possibility that the differences were due in some degree to age rather than nationality.

TABLE 2

SUICIDES AMONG LANGUAGE AND LITERATURE MAJORS VS. ALL OTHER SUBJECT MAJORS

Major subject group	Suicidal students		Total student body population	p
	n	%	%	
Language and literature	5	25	9	.012
All other majors	15	75	91	

Note.—Excludes three students who had not declared major subjects.

TABLE 3

GRADE-POINT AVERAGES FOR GRADUATE AND UNDERGRADUATE STUDENT SUICIDES

GPA	Suicidal students		Student population %	p
	n	%		
Class standing				
Undergraduate				
Above mean	10	91	50	.006
Below mean	1	09	50	
Graduate				
Above mean	4	40	50	*ns*
Below mean	6	60	50	

Note.—Excludes two students; one graduate, one undergraduate, who suicided during their first semester.

Major Subject

For this comparison, the suicidal subjects were divided into two categories, corresponding somewhat to William James' distinction between the "tough" and "tender minded." Of the 20 suicidal students who had declared majors, the breakdown was 10 students in the "tough-minded" or mechanical-mathematics group (Engineering, Professional, Physical Sciences, Biological Sciences, Agricultural majors) and 10 students in the "tender-minded" or esthetic-social group (Arts, Social Sciences, Language and Literature majors). Relative to their population proportions, there was a greater incidence of suicides in the tender-minded group, but not a large enough imbalance to achieve statistical significance. Further analysis, by individual subject groups, revealed that suicides were significantly more frequent among students majoring in languages and literature (five cases), especially English majors, who comprised three of the five cases (see Table 2).

TABLE 4

OBSERVED AND EXPECTED GPA OF STUDENT SUICIDES BY CLASS STANDING

Class standing	GPA	
	Observed	Expected
Undergraduate	3.18	2.50
Graduate	2.90	3.35

Grade-Point Average

Grade-point analysis required some basic adjustments since graduate and undergraduate grading systems are not directly comparable. In practice, an undergraduate "C" is approximately equivalent to a graduate "B." For the student population, the grade-point average (GPA) for undergraduates was 2.50, while for graduates it was 3.35 (calculated to the scale: A = 4, B = 3, C = 2, D = 1, F = 0). Given this discrepancy, it is obviously necessary to separately compare undergraduate and graduate students with reference to their respective grade-point distributions. When the suicidal students (excluding two who did not complete a full semester at UCB) are ranked by means of achievement above or below their population GPA, we find that two-thirds of them were above average while, by definition, only half of the general student body achieved this mark. Although suggestive of a tendency toward higher grades among suicidal students, the difference, in fact, did not achieve statistical significance. However, further analysis, distributing GPA by class standing, revealed a marked discrepancy between graduate and undergraduate students. This breakdown is detailed in Table 3 and reveals that of the 11 undergraduate students who committed suicide (after one complete semester at the University), 10 of them had surpassed the undergraduate GPA. For graduate student suicides, only 4 of the 10 who had completed a semester exceeded the graduate GPA. Despite the differential grading system that rewards the graduate student with more grade points for a similar level of work, the suicidal undergraduate students received a higher overall GPA than the graduate student suicides (see Table 4).

This finding seems to indicate that undergraduate and graduate suicides differ markedly from one another in terms of academic achievement. The undergraduate suicides performed on a level well above their fellow classmates and performed considerably better than did graduate suicides. Looking at the personal histories of these undergraduate students one discovers an interesting paradox. To an external observer, say someone viewing their transcripts, these students achieved spendidly in their academic pursuits. They had all been A or B students in high school since a B or better average is required for undergraduate admission, a policy which is estimated to limit entrance to the top 10–12% of graduating high school seniors. Reports from family and friends, however, reveal that self-satisfaction was not the case with these students. Rather, they seemed filled with doubts of their adequacy, dissatisfied with their grades, and despondent over their general academic aptitude. This exacerbated fear of failure was tempered somewhat by the fact that in every case of undergraduate suicide the final semester's GPA was lower ($\bar{x} = 2.53$) than the previous cumulative GPA ($\bar{x} = 3.34$). Another consideration is whether these students aspired to graduate school which requires a higher than average GPA (2.5–3.0 at UCB). Unfortunately, these exact data are not available; however, a check of those students in major subjects which definitely indicated future graduate work, for example, premedicine, revealed academic achievement in excess of grade requirements. Nevertheless, on balance, they were still achieving loftily above the average of their classmates. How can one explain their deep self-dissatisfaction despite contrary and objective indications of their competence? Two possible explanations suggest themselves: (*a*) The internal standards these students applied to themselves were so Olympian, the demands they imposed upon themselves so exacting, that they were destined to suffer frustration and disappointment no matter how well they fared; and/or (*b*) Whereas they had previously been crackerjack students in high school or junior college, excelling without much difficulty, the precipitous drop in grade points over the final semester threatened their feelings of self-esteem. Thus, faced by a sudden loss of status, they may have suicided as a response to this egoistic conflict. In any case, the discrepancy between perceived self-concept and objective reality indicates that a purely objective approach often obscures more than it reveals. What one needs to try and understand is the phenomenological response of the individual student. What is necessary to know is what inner standards, what idealized fantasy he uses to judge himself and his own personal worth. For the graduate student suicides as a group, there was no discrepancy between their academic achievements and what might be expected on the basis of the general population of graduate students. While they produced slightly below their population mean, the variation in this instance was primarily due to two students who were in considerable scholastic straits. Contrary to the undergraduates, graduate suicides showed no pattern of decline in their terminal semester GPA. Confirmation of the scholastic disparity between graduate and undergraduate suicides is further revealed by the irregular distribution of academic awards. Inspection of Table 5 indicates that undergraduate students garnered scholarship honors at a rate well beyond the general undergraduate population, while the graduate student suicides did not differ significantly from their classmates in earning academic awards. Even though graduate student awards were far more plentiful, the great majority of awards (10 of 11) were held by undergraduate student suicides.

Mental Health

Of the 23 student suicides, 8 had been referred to the student mental health service for psychiatric treatment (of the 8 students, apparently only 2 were diagnosed as psychotic reactions). These 8 cases comprised better than one-third of the student suicides, significantly exceeding the approximately 10% of the total student body population seen at the mental health facilities (see Table 1). Besides the 8 students known to the student psychiatric service, an additional 3 students were in private psychiatric treatment, making a total of almost 50% of the suicidal group who gave this particular indication of prior mental disturbance.

TABLE 5

SCHOLASTIC AWARDS BY CLASS STANDING

Class standing	Suicidal students		Student population	
	n	%	%	p
Undergraduate				
Scholarship	7	58	05	$<.001$
Nonscholarship	5	42	95	
Graduate				
Scholarship	1	10	23	*ns*
Nonscholarship	10	90	77	

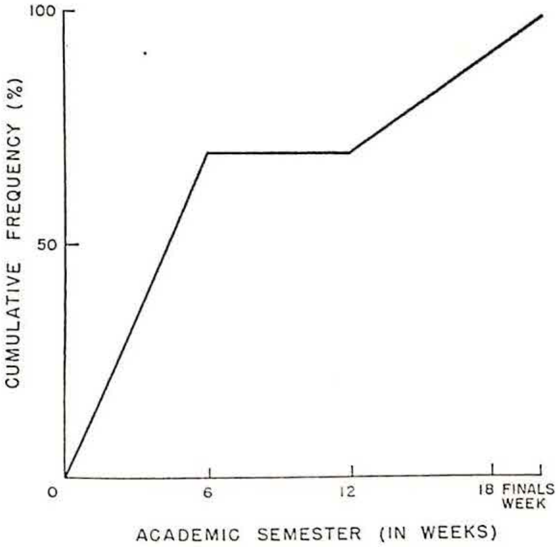


FIG. 2. Time distribution of student suicides, UCB, 1952-61.

Temporal-Situational Relationships

Among all causes of death, suicide allows for the greatest degree of volition. The suicidal person is in a position to choose the date, place, and method of his death, and it has long been speculated that there may be a special psychological significance to these choices. Through tracing the time, place, and method of student suicides, the following particular patterns were observed:

Time. When student suicides were charted by calendar months they formed a bimodal curve with peaks occurring during February and October. A more meaningful comparison obtained when the academic semester was used as the time interval. This distribution, as illustrated in Figure 2, challenges a frequently held belief about campus suicides. Academic folklore often explains student suicides as a response to the anxieties and

TABLE 6

METHODS OF SUICIDE USED BY UCB AND YALE STUDENTS

Method	UCB (1952–1961)		Yale (1920–1955)[a]	
	n	%	n	%
Firearms	8	35	10	40
Poisonings	6	26	3	12
Asphyxiation	4	17	5	20
Hanging	2	09	6	24
Jumping from high place	2	09	1	04
Cutting instruments	1	04	—	—
Total	23	100	25	100

[a] Source: Parrish, 1957, p. 589.

stresses of final examinations. Yet, surprisingly, the data showed that almost the reverse relationship held. Only 1 of the 23 student suicides was committed during finals. (Even that single instance may be dubiously related to final exams since this student was doing well in school and had expressed satisfaction with his "finals" performance.) Most of the suicides occurred at the beginning of the semester. When the semester is divided into three equivalent parts, the vast majority of cases, 16 out of 23, are found to occur during the first 6-week segment. (Actually, the period is only 5 weeks from when instruction begins; the first week is confined to registration procedures.) No cases were found during the second 6-week period which includes the mid-term examinations. Over the remaining third of the semester there were seven cases, just one of which occurred during finals week itself (always the last week of the semester). This irregular time distribution of student suicides departed significantly from uniform expectations ($x^2_2 = 16.8$, $p < .001$). Clearly, the old saw about suicides and finals was not supported. Instead, the danger period for student suicide was found to be the start, not the finish, of the school semester. Incidentally, the day of the week departed significantly from the null hypothesis of uniformity ($x^2_1 = 4.18$, $p < .05$) with almost one-half the cases occurring on Monday or Friday, terminals of the school week. Unfortunately, the data were none too precise since some cases were based on coroner's estimates as to the date of death.

The unexpectedly low correspondence between final examinations and the commission of student suicide bears some resemblance to a parallel phenomenon involving student mental health during the recent free speech activities on the UCB campus. In the course of these supposedly stressful times, there was a striking drop in admissions to the student mental health service (20% below average) and no recorded student suicides during the 1965 academic year. (Such behavior corresponds to the drop in suicides, psychosomatic illness, and neurotic conditions observed during both World Wars.) Why, in the midst of all the controversy, turmoil, and tempest was student mental health apparently enhanced? One possibility is that some students who had previously been grappling with internal problems now had the opportunity to act out, to ventilate their inner conflicts, and to displace their intrapunitive anger and hostility by redirecting it toward an external symbol, namely, the University. Perhaps it was the galvanized and heightened sense of community that facilitated mental well-being. Certainly many students felt involved in a common cause; probably, for some it imparted meaning to their lives where previously they had felt alienated and purposeless. If so, it was also a perfect antidote to the kinds of feelings that often drive people to self-destruction.

Place. Most of the students, 12 of 23, committed suicide at their residences. The next most frequent location was the University itself, upon whose grounds 4 students ended their lives. Three students were found dead in parked autos on isolated suburban roads. Another 3 suicided in out-of-town hotel rooms, and 1 student leaped from the San Francisco Golden Gate bridge. It is difficult to determine any significance to the site of these suicides, except for the 4 cases who killed themselves on the university grounds. Of these, the most symbolic suicide was the 1 student who jumped from the Campanile, an architectural landmark of the Berkeley campus.

Method. The most frequent agent of choice was firearms, followed by ingestions and asphyxiations. A comparison with the methods used by Yale student suicides (see Table 6) revealed considerable similarity in the methods employed by the two groups of students. The relatively larger number of poisonings among UCB students is most likely due to the more recent availability of tranquilizers and barbiturates.

For only two of the Berkeley cases was there the least equivocation about assigning suicide as the cause of death. These two cases, both involving ingestions of poisonous substances, were qualified as "probably suicide" but routinely coded as "deaths due to suicide." In at least 10 instances, suicide notes were left by the decedents. These notes ranged from simple instructions concerning the disposal of personal belongings to lengthy, literary dissertations, one of which finished by tersely quoting Camus: "Life as a human being is absurd."

Psychological Factors

A statistical approach, per se, can go just so far in describing the suicide-prone student. The additional use of case history materials provides a fuller, more clinically oriented dimension to the portrayal. As such, the following inferences were derived from anecdotal reports of friends and acquaintances of the students, along with those members of the University community whose lives they touched. From a preventive standpoint, the most pertinent questions which might be asked are, "What prodromal signs, what clues to suicide could be discerned from the personal lives of these students? Specifically, were there any indications or harbingers of their ultimate destinies?" Lastly, "Was there a characteristic conflict which precipitated their self-destructive actions?" The question of prodromal indications can be flatly answered "yes." There were numerous warnings in almost every case. At least five of the students had made past suicide attempts. Warnings of a more subtle nature could be discovered in the histories of the remaining students. For example, the pupil who went out of his way to modify an item on the medical history form. Where it had requested, "Whom shall we notify in case of emergency?" he crossed out the word "emergency" and substituted "death." Or the student who confided that he sometimes takes 10 or so nembutals because "I am an adventurer." Other students evidenced a long-standing infatuation with death, often initiating "bull sessions" about the futility of life, or making wry jokes about killing themselves. Prior to their suicides a disproportionately large number of these students were involved in psychiatric treatment. As a group, they presented similar symptomatic patterns featuring symptoms of insomnia, anorexia, and extreme moodiness, especially moods of despondency; in all, it was a psychological picture compatible with the general diagnosis of agitated depression.

Although their prodromal response to stress was very similar, the particular crises that precipitated their suicides were not. Bearing in mind that each individual case was unique, for purposes of description, the main prodromal conflicts could be classified into the following three categories:

1. *Concern over studies*—In many cases acquaintances of the students made such judgments as "he pushed himself too hard," "worried over grades," "felt his grades were not as good as he thought they should be," or similar scholastic anxieties which, they felt, triggered the suicidal crisis. It is difficult to evaluate these inferences since "worry over grades" is often seen by informants as a most likely explanation. At any event, if true, their exaggerated concern over studies contrasted vividly with generally excellent academic grades.

2. *Unusual physical complaints*—A number of the students compained of inability to eat or sleep, one student warranting a diagnosis of "avitaminosis." Others worried about possible deterioration such as the student who feared that his "failing sight" might ruin a prospective medical career. A few pupils, however, presented physical complaints of a bizarre semidelusional quality, for instance, the young man whose stomach literally persecuted him. From childhood on he had suffered from anorexia and "stomach ache." Although an exploratory laparotomy did not disclose anything, by the time he entered the University he was at least 50 pounds underweight, still wracked by chronic stomach pains. He then moved from his fraternity house, in the hope of gaining weight by selecting his own food. This plan proved to no avail, nor did extensive medical testing at the

student health service, all of which proved negative. He finally ended his torment, perhaps symbolically, by ingesting cyanide.

3. *Difficulties with interpersonal relationships*—Combined under this heading were two different types of conflicts, both reflecting problems in personal relationships. First were the students involved in stormy love affairs. Here the critical stresses were feelings of rejection which had been engendered by broken romances. In the one recorded instance of double suicide, the precipitating event was parental opposition to the youngsters' marriage. Much more typical, however, was the essentially asocial, withdrawn student. These particular students were uniformly described as terribly shy, virtually friendliness individuals, alienated from all but the most minimal social interactions. Frequently they had compensated for their personal solitude by increased study and almost total absorption in schoolwork. The most calamitous example of such human isolation was the student, dead for 18 days before he was found in his lonely room. It is a tragic commentary to his existence, and perhaps a cause for his suicide, that there were no friends, no people involved enough in his life to know, or to care, that he had been missing for well over 2 weeks.

Interpretation

Reviewing the results of the present study, one can reasonably conclude that significant associations between student suicide and numerous variables, both personal and environmental, have been demonstrated. Nonetheless, one cannot, with certitude, infer that these relationships are causal ones. This type of inference would require procedures more exacting than the limited epidemiological methods herein employed. For instance, the total student body population, used as a matched control or comparison group, included a number of students who had unsuccessfully attempted suicide. Quite possibly their inclusion diluted the significance of the obtained differences between suicidal and presumably nonsuicidal students. This is a relatively minor concern, compared to other more cautionary limitations. A primary concern is to what degree the observed relationships were spuriously increased by a common variable. For example, the correlation between student suicide and declining terminal GPA may very well be due to a third factor—emotional disturbance—which both depressed scholastic grades and led to self-destruction. As a corollary, it should be recognized that not all of the selected variables were independent of one another. It is known for one that age and class standing are highly dependent, and it was observed, also, that the variable of age probably confounded to some degree the comparisons by sex and by nationality. Another area of uncertainty concerns the time-order sequence of student suicide. One is unable to state, with certainty, which comes first, the disturbed student or the stresses of student life. Are the suicides due to selection into colleges of mentally unstable individuals or are they due to competitive pressures of the academic environment? The fullest answer to these questions will only come from further research. Toward this goal some salient lines of inquiry could include: the investigation of student suicide attempters and student accident cases, postcollegiate follow-up studies, and the use of "psychological autopsy" procedures, as described by Shneidman and Farberow (1961).

Within the expressed limits of the study design, what predictions about the future suicide problem are warranted? Extrapolating from results of the present study, it appears that a future increase of student suicides may be expected. This increase should occur as a function of two variables, that is, age and academic competition, both of which are directly correlated to student suicides, and both of which are slated to increase in future student body populations. Average student age is already rising as a result of ever increasing proportions of graduate students in the American university system. For example, architects of the UCB educational master plan are considering an ultimate 50:50 graduate-undergraduate ratio. The second variable, academic competition, will likely increase as a result of mounting public demands for quasi-universal college education. As a case in point, the enrollment demands at UCB have already exceeded the available academic supply. Consequently, it has been necessary to restrict enrollment to the upper-most fraction of high school graduating classes. If accepted, the pressure on the student to achieve and maintain very high GPAs gives no indication of abatement. In fact, the situation ominously resembles a suicidal problem which prevails among the youth of Japan. In the Japanese case there are tremendous pressures to attend college, and those students who fail to gain entrance frequently turn to suicide as a solution to their dilemmas. Such conflicts, in addition to a more accepting cultural attitude, have probably helped to make Japan "a country of youthful suicides where suicide has become the number one cause of death in individuals under 30 [DeVos, 1964, p. 6]."

SUMMARY

The purpose of this study was to identify distinctive attributes of the suicidal student, and to determine those environmental conditions which heighten his susceptibility to suicide.

Using an epidemiological approach, demographic comparisons were made between the sample of 23 UCB students who committed suicide during the years 1952–1961 inclusive, and the total student body population for those years. As an additional procedure, the temporal-situational characteristics of student suicides were described and analyzed.

The main findings of the research were:

1. Suicidal students could be significantly differentiated from their classmates on the variables of age, class standing, major subject, nationality, emotional condition, and academic achievement. Compared to the student population at large, the suicidal group was older, contained greater proportions of graduates, language majors, and foreign students, and gave more indications of emotional disturbance. In addition, the undergraduate suicides fared much better than their fellow students in matters of academic achievement.

2. Contrary to the popular belief that suicides frequently occur during final examinations week, time relationships indicated that the peak danger period for student suicides was the beginning (first 6 weeks), not the midterm, nor end of the semester.

3. Most of the students gave recurrent warnings of their suicidal intent. Many of them presented a similar prodromal pattern marked by anorexia, insomnia, and periods of despondency.

4. Major precipitating factors were: Worry over schoolwork, chronic concerns about physical health (sometimes of a decidedly bizarre nature), and difficulties with interpersonal relationships. This last category contained some students who had reacted to romantic rejections but, for the most part, comprised the emotionally withdrawn and socially isolated student.

5. A future increase of student suicides was predicted on the basis of changes taking place in the age structure of college populations and in the competitive pressures of student life.

REFERENCES

BRAATEN, J., & DARLING, C. Suicidal tendencies among college students. *Psychiatric Quarterly,* 1962, **36**, 665–692.

BRUYN, H. B., & SEIDEN, R. H. Student suicide: Fact or fancy? *Journal of the American College Health Association,* 1965, **14**, 69–77.

DEVOS, G. Role narcissism and the etiology of Japanese suicide. Berkeley, Calif.: Institute of International Studies, University of California, 1964. (Mimeo)

DUBLIN, L. I. *Suicide: A sociological and statistical study.* New York: Ronald, 1963.

GIBBS, J. P., & MARTIN, W. T. *Status integration and suicide.* Eugene: Oregon University Press, 1964.

HILL, A. B. *Principles of medical statistics.* New York: Oxford University Press, 1961.

JENSEN, V. W. Evaluating the suicidal impulse in the university setting. *Journal Lancet,* 1955, **75**, 441–444.

MACMAHON, B., PUGH, T. F., & IPSEN, J. *Epidemiological methods.* Boston: Little, Brown, 1960.

PARRISH, H. M. Epidemiology of suicide among college students. *Yale Journal of Biology and Medicine,* 1957, **29**, 585–595.

RAPHAEL, T., POWER, S. H., & BERRIDGE, W. L. The question of suicide as a problem in college mental hygiene. *American Journal of Orthopsychiatry,* 1937, **7**, 1–14.

SHNEIDMAN, E. S., & FARBEROW, N. L. Sample investigations of equivocal deaths. In N. L. Farberow & E. S. Shneidman (Eds.), *The cry for help.* New York: McGraw-Hill, 1961. Pp. 118–128.

SUSLOW, S. A *ten-year survey of certain demographic characteristics of the student population.* Berkeley: Office of the Registrar, University of California, 1963. (Mimeo)

TEMBY, W. D. Suicide. In G. B. Blaine & C. G. McArthur (Eds.), *Emotional problems of the student.* New York: Appleton-Century-Crofts, 1961. Pp. 133–152.

(Received November 17, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 400–406

AN EXPERIMENTAL DEMONSTRATION OF HYPNOSIS AS ROLE ENACTMENT[1]

WILLIAM C. COE

Langley Porter Neuropsychiatric Institute, San Francisco

AND

THEODORE R. SARBIN

University of California, Berkeley

3 variables drawn from role theory were investigated as predictors of hypnotic responsiveness: congruence of self and role (assessed by questionnaire); role expectations (assessed by questionnaire); and role-taking aptitude (drama students vs. science students). 168 Ss (78 drama students, 91 science students) were administered the 2 questionnaires and the Harvard Group Scale of Hypnotic Susceptibility. 2 hypotheses were tested. (1) Ss high on any 1 of the variables would show more hypnotic responsiveness than Ss low on the same variable. (2) Ss high on all 3 variables would show more hypnotic responsiveness than those high on 2 variables; Ss high on 2 variables would show more hypnotic responsiveness than those high on 1 variable; and so on. Hypothesis 1 was confirmed for role-expectation and role-taking aptitude but not for congruence of self and role. Hypothesis 2 was confirmed. Qualifications of the variables and implications for further research were discussed.

This article examines the utility of a recent revision of role theory to account for hypnotic behavior (Sarbin, 1964). The utility of a set of metaphors drawn from dramaturgical sources to describe social psychological behavior has received considerable support (e.g., Goffman, 1961; Mead, 1934; Sarbin, 1950, 1952, 1964). Social interaction is regarded as the enactment of roles, given certain antecedent and concurrent conditions. Since hypnosis is a social psychological phenomenon, that is, is dependent upon communication and reciprocal action, it follows that it can be profitably studied with the aid of role-theory concepts. The performance of a person who serves as the subject (*S*) in the hypnotic situation is subsumed under the rubric *role enactment*. The question to which efforts are addressed is: What are the conditions that influence the accuracy of hypnotic role enactment? (Accuracy may also be read as validity, convincingness, or effectiveness.)

The dependent variable, accuracy of hypnotic role enactment, was operationally defined as *S*'s score on the Harvard Group Scale of Hypnotic Susceptibility (Shor & Orne, 1962). This scale is a modification, for group administration, of the Stanford Hypnotic Susceptibility Sale, Form A (Weitzenhoffer & Hilgard, 1959). The experimenter (*E*) reads a standard hypnotic induction procedure. A waking suggestion is given first. It is followed by suggestions of relaxation and visual fixation aimed at eye closure. A number of hypnotic tasks are then suggested to the group, including and terminating with, the suggestions for posthypnotic amnesia and the performance of a posthypnotic act on cue. After the hypnotic period, *S*s answered a group hypnotic scale that scores their responses on the basis of objective criteria, that is, motor responses that would be visible to an observer. A separate section of this is provided for subjective experiences. The group hypnotic scale has been found to have a score distribution similar to the individually administered scale, and a validity coefficient of .74 with independent ratings of hypnotic depth (Shor & Orne, 1963). Self-scoring correlated .74 with the observer scoring in one sample and .81 in another sample (Bentler & Hilgard, 1963), almost as high as the reliability reported for the individually administered Form A. Bentler and Roberts (1963) also found the group scale to correlate .72 and .67 with the Stanford Form C hypnotic scale. These findings indicate that this scale as a measure of hypnotic responsiveness is comparable to the individually administered standard scales of hypnotic susceptibility.

---

[1] This research was supported in part by National Institute of Mental Health predoctoral research Grant 1FIMH-19,838-01. The questionnaires used in this study have been deposited with the American Documentation Institute. Order Document No. 9182 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Three independent variables derived from role theoretical concepts were operationally defined and measured.

I. Congruence of Self and Role: If the self-characteristics of a person are compatible with the requirements of the prospective role, the person has a potential for enacting that role. There are two aspects that would seem required of an *S* if he is to find the hypnotic role congruent with his self-system. One is that he accept the possibility of altered states of attention, new and different experiences, and, from the popular stereotype of hypnosis, mystical occurrences. The second requirement is that *S* see himself as one who can concentrate and maintain his attention on specific tasks, the hypnotic instructions in this case. The Congruence Questionnaire was developed to measure *S*'s degree of congruency with the requirements of the hypnotic role. It was composed of two scales with items thought to tap each of the previously mentioned requirements. The Acceptance of Altered Processes Scale contained items such as:

I would find it interesting to see how my world would change with altered perceptions of it. (True)
Would you like to get beyond the world of logic and reason and experience something new and different? (True) (Yes)

The Role Absorption Scale contained items such as:

I usually have to stop and think before I act even in trifling matters. (False)
I sometimes find that when I'm studying hard I don't notice the passage of time. (True)

The total score on this 50-item questionnaire was *S*'s congruence measure. Those *S*s scoring above the sample mean were arbitrarily considered high-congruence *S*s, those scoring below the mean were low-congruence *S*s.

II. Role Expectation: The probability of enacting a role validity is increased if a person knows what behaviors are expected in a role. An Expectation Questionnaire was constructed to measure *S*'s accuracy of expectations for the hypnotic role. Two scales made up this 30-item questionnaire. The first, the General Expectation Scale, measured *S*'s general expectations of hypnotic role enactment. Examples of these items are:

Women who are hypnotized before childbirth often suffer a good deal of pain anyway. (False)
I believe I can be cured of nearly any kind of psychosomatic illness through the use of hypnosis. (True)

The Specific Expectations Scale measured the accuracy of an *S*'s expectations for the specific role he was to enact, the Harvard Group Scale of Hypnotic Susceptibility. If he answered in the keyed direction, he would be scored as a good hypnotic *S* in the experimental session. Examples of these items were:

The first signs of hypnosis are often slight changes in vision. (True)
Vivid thoughts brought about under hypnosis seem real to the hypnotized person. (True)

As with the Congruence Questionnaire, *S*s scoring above the sample mean were considered high-expectation *S*s, those scoring below the mean as low-expectation *S*s.

III. Role-Playing Aptitude: Dramatic arts students were selected as high role-playing-aptitude *S*s. Presumably, they are flexible in taking on new roles and have been rewarded for following a director's suggestions. Also, and what seems more important, they have practiced motor and imagination behaviors specifically required in the hypnotic role. Students majoring in science were considered to be low role-playing-aptitude *S*s because their present goals presumably require little practice in changing roles. Their training attempts to reach an objective approach in their commerce with events and objects so that subjectivity and introspective practice are minimized.

Three additional variables contributing to effective role enactment (Sarbin, 1964) were controlled. (*a*) Role perception, the perceived

TABLE 1

COMPARISON DATA BETWEEN THE DRAMATIC ARTS SUBJECTS AND THE SCIENCE SUBJECTS

	Drama arts majors[a]	Science majors[b]
Hypnotized before	10%	7%
Age range	17–49	17–30
M age	21.40	22.03

[a] $n = 34$ males and 43 females.
[b] $n = 49$ males and 42 females.

status of the other, was unambiguous. That is, *E* announced he was the hypnotist. (*b*) Reinforcement of the ecology was similar for all *S*s because the hypnotic procedure was a verbatim administration with no differential reinforcing properties. (*c*) Role demands cannot be as easily assumed controlled because of their implicit nature; however, it was felt that differences in *S*s' perceptions of their task would occur randomly throughout the groups.

Two hypotheses followed from role theoretical predictions.

1. The *S*s high on one of the measured variables would be better hypnotic *S*s than those low on the same variable.

2. The *S*s high on more measured variables would be better hypnotic *S*s than those high on fewer variables. That is, those *S*s high on all three variables would score higher on the hypnotic scale than those high on two variables, and so on.

METHOD

Subjects. The total sample was made up of 168 students who had completed 2 years or more of college work. There were 78 *S*s majoring in dramatic arts and 91 *S*s majoring in the biological and physical sciences. Table 1 presents comparison data for the drama majors and science majors. The *S*s were solicited in their classrooms. The *E* announced that he was seeking volunteers for a psychology experiment and that anyone participating would be paid \$2.00 for 1½ hours. After the announcement, sign-up sheets for several scheduled experimental sessions were distributed. There was no mention of the nature of the experiment. The *E* collected the sign-up sheets before the class period ended. Approximately 45% of the *S*s addressed volunteered for the experiment.

Procedure. The *S*s reported to the experimental room (a typical classroom with a capacity for 35 *S*s) at the time prescribed on the sign-up sheets. The *E* announced that the first part of the experiment would be two short questionnaires. The *S*s were instructed to answer quickly and if in doubt about an item to respond with their first thought. The Congruence Questionnaire was administered first and the Expectation Questionnaire second. No more than 30 minutes was allowed for the two questionnaires. The group was then informed that a standard hypnotic scale would be administered and that all of them would probably be hypnotized to some degree. The *S*s were told they were free to leave at that time if they wished. Only two *S*s chose to leave (two female drama students in the final scheduled session). The Harvard Group Scale of Hypnotic Susceptibility, Form A (Shor & Orne, 1962) was administered. The *S*s completed the self-scoring booklets, handed in their materials, and were paid. None of the *S*s required more than 90 minutes for the entire session. Before anyone left the session, *S*s were asked not to reveal what had happened until after the last experimental session. None of the *S*s had to conceal this information for more than 4 days. Most of the *S*s did not have a meeting of the class in which they had been contacted between the experimental session they attended and the last experimental session scheduled for that class. Although *E* questioned *S*s in the later sessions, none indicated he had expected to be hypnotized. The general group response to the announcement of hypnosis was a murmur of surprise and an increase in group seriousness.

Nine groups of *S*s comprised the experimental sample. The group sizes ranged from 7 to 30. The dramatic arts *S*s and science *S*s were not seen in the same experimental sessions.

Experimental design. A three-factor (two levels per factor) analysis-of-variance model was used to test the three variables. The two levels (highs and lows) of the congruence-of-self-and-role variable and the role-expectation variable were arbitrarily defined by selecting those *S*s above the sample mean of each questionnaire as high-congruent and/or high-expectation *S*s. Those *S*s below the mean were selected as low-congruent and/or low-expectation *S*s. The dramatic arts students were considered high role-taking-aptitude *S*s; the science students low role-taking-aptitude *S*s. As a result of this procedure each *S* received a classification on the three variables depending on his two questionnaire scores and his college major.

RESULTS[2]

Because of unequal cell frequencies and nonindependent variables, the least-squares analysis-of-variance solution suggested by

---

[2] Although this study was not designed to investigate sex differences in hypnotic responsiveness, mention should be made of two possible interactions between sex and role-theory variables. The first, an interaction between sex and role-taking aptitude (graphically shown in Coe, 1964), suggests that while females, either drama or science majors, are

Winer (1962) was employed.[3] The analysis-of-variance source table is reported in Table 2. Two role-theory variables were significant: Expectation ($p < .05$) and Aptitude ($p < .001$). The Congruence variable and the interaction terms were not significant.[4]

TABLE 3

3 *t*-TEST RESULTS BETWEEN THE 2 LEVELS OF EACH ROLE-THEORY VARIABLE

Role-theory variable	*N*	*M*	σ	*t*
Congruence				
High *S*s	80	7.23	2.82	2.24*
Low *S*s	88	6.21	3.04	
Expectation				
High *S*s	68	7.37	2.76	2.47*
Low *S*s	100	6.23	3.04	
Aptitude				
High *S*s	77	7.98	2.51	5.60**
Low *S*s	91	5.61	2.91	

* $p < .05$.
** $p < .01$.

Table 3 shows the *t*-test results between the two levels of each variable. The *S*s classified as high on each of the variables are significantly higher hypnotic scorers than *S*s classified as low: Congruence ($p < .05$), Expectation ($p < .05$), Aptitude ($p < .01$). The Congruence variable significantly differentiates if it is tested by itself; however, it loses its significance in the analysis-of-variance least-squares solution. This analysis in summary shows that *S*s high on Expectation and/or Aptitude are more responsive hypnotic *S*s. The *S*s classified as high-congruent *S*s have significantly higher hypnotic scores than low-congruent *S*s only when the other two variables are not taken into account.

The *S*s were grouped according to the number of role-theory variables on which they were classified as high: 28 *S*s were high

TABLE 2

SOURCE TABLE FOR LEAST-SQUARES SOLUTION OF 3 FACTORS ($2 \times 2 \times 2$) ANALYSIS OF VARIANCE: 3 ROLE-THEORY VARIABLES

Source	*SS*	*df*	*MS*	*F*
Congruence	.11	1	.11	.02
Expectation	33.56	1	33.56	4.52*
Aptitude	124.51	1	124.51	16.78**
Congruence × Expectation	.30	1	.30	.04
Congruence × Aptitude	11.89	1	11.89	1.60
Expectation × Aptitude	2.28	1	2.28	.31
Three-way interaction	0.00	1	0.00	.00
Within (Error term)	1187.20	160	7.42	

* $p < .05$.
** $p < .01$.

---

more responsive to hypnosis than are the males in the same major, the drama females outdo their male counterparts more than the science females, that is, being a female in the drama group correlated .30 ($p < .01$) with total hypnotic score while the same coefficient in the science group was only .17 ($p < .05$). Another possible interaction appears between sex and the congruence variable. Women tend to get higher congruence scores ($r = .16$, $p < .05$); however, this relationship appears primarily due to the difference between science females and science males. In the science group females tend to score higher on the variable ($r = .23$, $p < .05$) while this is not true in the drama group ($r = .02$, $p > .4$). Although these relationships are not particularly high, they do suggest that sex is a variable which may affect results and should either be controlled or investigated in hypnotic research.

Thus, in the present study, since the analysis-of-variance groups contained different numbers of males and females, the mean difference between the male and female hypnotic score in the total sample (1.54) was added to each male score before the analysis-of-variance computation. This procedure is essentially the same as an analysis of covariance (Professor William Meredith, personal communication). Technical difficulties at the time prevented the use of the covariance option in the computer program.

[3] Complete statistical rational for the least-squares solution may be obtained in BIMD 1090 Computer Programs, University of California at Los Angeles Medical Center, Los Angeles, California. The program was BIMD 14.

[4] The questionnaires were not "constructed" to control acquiescence response set. The Congruence Questionnaire key favored "true" responses (70%) as did the key for the Expectation Questionnaire (66.7%). In order to examine the possibility that "yea saying" could account for higher questionnaire scores in the higher hypnotic subjects, a frequency count of the items in the keyed direction was made for two groups: those scoring seven or higher on the group scale, and those scoring less than seven. These two groups did not differ significantly on the percentage of "true" responses of the total responses in the keyed direction. On the Congruence Questionnaire the high hypnotic group's total keyed score was based on 75% true responses while the low hypnotic group was 73.1%. The high group showed 68.6% as opposed to the low group's 68.4% true from the total keyed responses on the Expectation Questionnaire. These findings suggest that there may have been a tendency to answer the true items more readily, but this tendency does not account for the difference between high and low hypnotic scores.

TABLE 4

t-TEST RESULTS BETWEEN GROUPS OF SUBJECTS BASED ON THE NUMBER OF VARIABLES ON WHICH THEY WERE CLASSIFIED AS HIGH

Comparison groups	*M*	σ	*df*	*t*
High on three variables	8.48	2.33	68	2.97*
versus				
High on two variables	7.32	2.51		
High on two variables	7.32	2.51	95	2.91*
versus				
High on one variable	6.36	2.72		
High on one variable	6.36	2.72	94	3.03*
versus				
Low on all variables	5.30	2.95		

* $p < .01$.

on all three, 44 *S*s were high on two, 53 *S*s were high on one, 43 *S*s were low on all. A highly significant *F* ($p < .001$) resulted from a least-squares simple analysis-of-variance solution computed among the four groups.

Table 4 presents the *t*-test results among the four groups. Any group of *S*s that is higher on more role-theory variables than another group scores significantly higher on the group hypnotic scale ($p < .01$).

DISCUSSION

The results support the notion that hypnotic responsiveness may be studied profitably if it is considered as role enactment. Some specifications of the variables contributing to the accuracy of hypnotic role enactment may also be drawn from the results.

The role-playing-aptitude variable is the most discriminating of the three variables studied. The correlation between aptitude and hypnotic score was $r = .41$ ($p < .01$). We emphasize the assumption behind our specification of aptitude. Dramatic students were considered high hypnotic role-taking-aptitude *S*s because they are generally flexible in taking new roles, are motivated to take directions from others, and have practiced attending to their gestural and other motoric behaviors and to internal stimuli. Being able to perform tasks similar to those required in hypnosis increases the convincingness of their acting performances (see, e.g., Stanislavski, 1961). In role-theory terms, these attributes are conceptualized as general and role-specific aptitudes. The general aspect is role flexibility or changeability; the specific aspects are the motoric and cognitive behaviors required in the hypnotic role.

The role-expectation variable was somewhat less discriminating than aptitude. The correlation between expectation and hypnosis was $r = .17$ ($p < .05$). Knowing what behaviors are expected in the hypnotic role aids performance; however, this knowledge seems to help primarily in the central range of hypnotic responsiveness (scores of 4.0–8.0). This qualification of the expectation variable was discovered by ranking *S*s on their hypnotic scores, then testing between the high and low 50% groups and the high and low 25% groups. The upper 50% group scored significantly higher on the Expectation Questionnaire than the lower 50% group ($t = 3.06$, $p < .01$); however, there was not a significant difference in expectation scores between the upper and lower 25% groups ($t = 1.94$, $p < .1$). It seems that the central hypnotic range contains *S*s who differ significantly on accuracy of expectations. Apparently, if one is high or low in role-taking aptitude, the knowledge of the role expectations has little effect on the convincingness of role enactment. Those *S*s in the medium role-taking-aptitude range are able to enact the role with greater accuracy if they hold the proper role expectations. From this qualifying hypothesis one would expect the Expectation Questionnaire to differentiate a sample of social science students (assumed medium role-taking-aptitude *S*s) on hypnotic responsiveness to a greater degree than the present sample. This conclusion is amendable to empirical test.

Congruence of self and role was not a discriminating variable in the present design. This appears paradoxical because other researchers using questionnaires of similar content have reported significant correlations with hypnotic susceptibility (Aas, 1962; Anderson, 1963; Shor, Orne, & O'Connell, 1962). The Congruence Questionnaire correlated $r = .25$ ($p < .01$) with the Group Hypnotic score and $r = .44$ ($p < .01$) with the role-taking-aptitude variable. What apparently occurred in the present study was that Congruence functioned as a selection variable for high role-taking-aptitude *S*s; thus, by itself it did not contribute in the selection of high hypnotic scorers. That is, the aptitude variable accounted for the hypnotic variance so that in a multiple-regression model the congruence variable did not receive significant weighting. In retrospect, one finds an indication that this might have been predicted. Aas and Lauer (1962) factor-analyzed 23 cross-validated items from their Experience Inventory (items similar to those on the Congruence Questionnaire) with items from the Stanford Hypnotic Susceptibility Scales, Form A and C. Their first factor was primarily a hypnotic performance factor; the second a questionnaire factor. Only one item from their Experience Inventory loaded above .30 on the hypnotic factor and less than .20 on the inventory factor. This one item was "Having experienced becoming the character when acting in a play." It is apparent that this item selected *S*s who had been interested in dramatics or were in dramatics, that is, high role-taking-aptitude *S*s. The conclusion is that congruence of self and role will predict hypnotic susceptibility; however, its predictive power lies in its ability to select *S*s with high role-taking aptitudes.

The important place of role-taking aptitude in hypnotic behavior suggested in this study deserves further research. Sarbin and Lim (1963) found that *S*s (not drama students) who were rated by drama teachers as good, improvising, role takers were better hypnotic *S*s than *S*s rated low on this task. The importance of role-taking aptitude points to the need for more intensive investigation of individuals who are high and low in these abilities. Isolating important differences between these *S*s may provide fruitful insights into hypnotic behavior.

It is recognized that homogeneous grouping of drama *S*s and science *S*s could present different demand characteristics for the experimental groups. For example, the drama students might infer that their ability to concentrate was being tested and thus respond more readily to the instructions. While this does not invalidate the experimental results, it may weaken the theoretical conclusions. That is, the difference between drama *S*s and science *S*s may be due to additional antecedent or concurrent variables. Replications of this study should mix the two groups in the experimental sessions in order to eliminate this doubt.

That the findings are consistent with other theoretical formulations cannot be gainsaid. For example, Arnold (1946) would probably have predicted the correlation between hypnotic role enactment and role-taking aptitude through the mediation of skill in "imagining." Hilgard (1964), too, might have made the same prediction from his developmental interactive theory of hypnosis—drama students and science students presumably are products of different kinds of socialization. These theories are not contrary nor contradictory to role theory. The point has been made elsewhere (Sarbin, 1943, 1950) that skill in role taking is related to the ability to use as-if formulations (imaginings) and that such skill is a function of a particular kind of reactional biography. One might also add that the variables employed in this study are more likely to emerge from a dramaturgical model than from a purely cognitive or dynamic model.

REFERENCES

AAS, A. Non-hypnotic experiences related to hypnotizability in male and female college students. *Scandinavian Journal of Psychology,* 1962, 3, 112–121.

AAS, A., & LAUER, L. A factor analytic study of hypnotizability and related personal experiences. *International Journal of Clinical and Experimental Hypnosis,* 1962, **10**, 181–199.

ANDERSON, M. L. Correlates of hypnotic performance: An historical and role-theoretical analysis. Unpublished doctoral dissertation, University of California, Berkeley, 1963.

ARNOLD, M. On the mechanism of suggestion and hypnosis. *Journal of Abnormal and Social Psychology,* 1946, **41**, 107–128.

BENTLER, P. M., & HILGARD, E. R. A comparison of group and individual induction of hypnosis with self-scoring and observer-score. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11**, 49–54.

BENTLER, P. M., & ROBERTS, M. R. Hypnotic susceptibility assessed in large groups. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11**, 93–97.

COE, W. C. Further norms on the Harvard Group Scale of Hypnotic Susceptibility, Form A. *International Journal of Clinical and Experimental Hypnosis,* 1964, **12**, 184–190.

Goffman, E. *Asylums.* Chicago: Aldine, 1961.

Hilgard, E. R. The motivational relevance of hypnosis. In D. Levine (Ed.), *Nebraska symposium on motivation: 1964.* Lincoln: University of Nebraska Press, 1961. Pp. 1–44.

Mead, G. H. *Mind, self, and society.* Chicago: University of Chicago Press, 1934.

Sarbin, T. R. The concept of role-taking. *Sociometry,* 1943, **6,** 273–285.

Sarbin, T. R. Contributions to role-taking theory: I. Hypnotic behavior. *Psychological Review,* 1950, **57,** 255–270.

Sarbin, T. R. Contributions to role-taking theory: III. A preface to a psychological analysis of self. *Psychological Review,* 1952, **59,** 11–22.

Sarbin, T. R. Role theoretical interpretation of psychological change. In P. Werchel & D. Byrne (Eds.), *Psychological change.* New York: Wiley, 1964. Pp. 176–219.

Sarbin, T. R., & Lim, D. T. Some evidence in support of the role-taking hypothesis in hypnosis. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11,** 98–103.

Shor, R. E., & Orne, E. C. *The Harvard Group Scale of Hypnotic Susceptibility, Form A. An adaptation for group administration with self-report scoring of the Stanford Hypnotic Susceptibility Scale, Form A.* Palo Alto, Calif.: Consulting Psychologists Press, 1962.

Shor, R. E., & Orne, E. C. Norms on the Harvard Group Scale of Hypnotic Susceptibility, Form A. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11,** 39–48.

Shor, R. E., Orne, M. T., & O'Connell, D. Validation and cross-validation of a scale of self-reported personal experiences which predicts hypnotizability. *Journal of Psychology,* 1962, **53,** 55–75.

Stanislavski, C. *An actor prepares.* New York: Theatre Arts Books, 1961.

Weitzenhoffer, A. M., & Hilgard, E. R. *Stanford Hypnotic Susceptibility Scale, Forms A and B.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received August 6, 1964)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 407–412

CONFORMITY AND ACHIEVEMENT-RELATED CHARACTERISTICS OF DEPRESSED PATIENTS[1]

EDWARD S. KATKIN, DOROTHY BLUM SASMOR
State University of New York at Buffalo

AND

RAMON TAN
E. J. Meyer Memorial Hospital, Buffalo

Conformity and achievement-related characteristics of 10 hospitalized depressed patients and 11 matched acute schizophrenics were studied in an Asch-type conformity situation, and by the use of self-report inventories. It was found that in the conformity situation, depressed Ss showed a greater tendency to conform to social pressure than did the controls. When conformity and achievement were assessed by the use of self-reports, the results were either nonsignificant or opposite to those found in the conformity situation. The results were compared with recent theory and research dealing with cyclical affective disorders.

The goal of the present study was to extend recent theory and research concerning manic-depressive disorders to depressed patients. On the basis of information obtained from intensive psychoanalysis of 12 manic-depressives, Cohen, Baker, Cohen, Fromm-Reichmann, and Weigert (1954) concluded that these patients shared common characteristics of excessive conformity, dependency, and need for social approval. Furthermore, Cohen et al. suggested that manic-depressives were frequently used by their families as "esteem-getting" instruments; consequently they developed a sense of the importance of achievement as an instrumental act, without incorporating an internalized need for achievement. A follow-up study by Gibson, Cohen, and Cohen (1959) confirmed the findings of Cohen et al. and suggested that manic-depressive patients were characterized as possessing an excessive need for social approval, particularly from their parents. These findings were consistent with earlier work (Finley & Wilson, 1951) which suggested that manic-depressives have a strong need for social approval and show strong adherence to traditional family ideology.

Recent experimental tests of Cohen et al.'s hypotheses within the framework of need-achievement theory (McClelland, Atkinson, Clark, & Lowell, 1953) have generally obtained positive results. For instance, Becker (1960) found that manic-depressives in remission demonstrated characteristics associated with high "value achievement"—that is, they valued achievement more for the approval it obtained from others than for its own sake. In addition, Becker reported that remitted manic-depressives showed strong conformity tendencies and subscribed to authoritarian values, especially those pertaining to traditional family ideology. Further evidence supporting Becker's hypotheses has been reported for a sample of manic-depressives (Spielberger, Parker, & Becker, 1963) as well as for a sample of nonpsychiatric cyclothymic college students (Becker & Nichols, 1964).

Actuarially, the number of patients diagnosed manic-depressive has declined in recent years, and it has been suggested that this diagnostic category accounts for no more than 3% of present hospital admissions (Malzberg,

---

[1] This study was supported in part by a grant-in-aid (50-591-A) from the Research Foundation of the State University of New York. The authors are indebted to David Cavenaugh and W. J. Stryjewski of Buffalo State Hospital for their cooperation in carrying out this study and to J. S. Efran for his critical review of the manuscript. This study is based in part upon a thesis by the second author (Blum, 1966) submitted to the Graduate School of Arts and Sciences at the State University of New York at Buffalo in partial fulfillment of the requirements for the MA degree.

1959). However, according to Arieti (1959), clinical literature indicates that "patients previously diagnosed as manic-depressive may now be diagnosed as suffering from involutional psychosis [p. 441]"; furthermore, Arieti has suggested that among the few remaining manic-depressives "we may observe that the attacks of depression outnumber those of mania. Some patients may undergo a conspicuous number of depressions without ever having a manic attack [p. 429]." These reports, as well as Cohen et al.'s (1954) contention that "the basic psychotic pattern [of manic-depressives] is the depressive one [p. 225]," suggest the fruitfulness of extending recent research with manic-depressives to a sample of depressed patients.

Thus the major purpose of the present study was to investigate conformity and achievement-related characteristics of hospitalized depressed patients. A second purpose was to develop a technique for the assessment of conformity behavior, since previous studies with manic-depressives utilized no index of actual conformity in a specified situation, but rather inferred conformity from self-reports. The major hypothesis of the study was that depressed patients would show more conformity behavior than a control group. In addition it was hypothesized that depressed patients would show higher scores than a control group on (*a*) a value-achievement scale, (*b*) the California F(ascism) scale, (*c*) a measure of traditional family ideology, and (*d*) a measure of "social desirability."

METHOD

Subjects

In previous work with manic-depressive subjects (*S*s) the experimental group consisted of patients in remission; therefore, control groups usually consisted of nonpsychiatric *S*s recruited from the community (see Becker, 1958, 1960). In the present study, however, hospitalized depressed patients were employed for the experimental group. Thus, it seemed imperative to utilize hospitalized patients as a control group. Furthermore, in order to be sure that any observed differences could be attributed to unique characteristics of depressed patients rather than psychiatric patients in general, it was decided to utilize a control group consisting of hospitalized psychiatric patients. It was felt that acute schizophrenics, preferably of the paranoid type—since they frequently have dysphoric affect, but presumably different personality dynamics from depressives—would be appropriate *S*s for the control group.

Only female patients were selected for the study, and only if the duration of their present hospitalization was shorter than 6 months, and if they had no record of electroconvulsive therapy (ECT) within the past 5 years. The *S*s were assigned to the experimental group if they were diagnosed as involutional depression or psychotic depressive reaction, and if they had no history of schizophrenia; they were assigned to the control group if they were diagnosed as schizophrenic, and if they had no history of depressive disorder. As an additional screening device, no *S* was assigned to the depressed (D) group if she had a *T* score below 70 on the MMPI *D* scale, and no *S* was assigned to the schizophrenia (Sx) group if she had a *T* score above 60 on the MMPI *D* scale.[2] In this manner 10 *S*s were selected for the D group; their ages ranged from 25 to 56, with a mean of 40.8 and a standard deviation of 10.7. Eleven *S*s were selected for the Sx group; their ages ranged from 29 to 54, with a mean of 41.8 and a standard deviation of 7.1. Eight *S*s in the Sx group had been diagnosed paranoid schizophrenia.

Procedure

All *S*s for the study were originally selected for psychometric screening by a clinical psychiatrist on the basis of clinical diagnosis of depression or schizophrenia, length of hospitalization, and history of ECT. The *S*s were then administered, individually, by a female experimenter (*E*), a modified form of the MMPI which included the standard MMPI clinical and validity scales, as well as the Edwards Social Desirability (ESD) scale (Edwards, 1957). In addition, all *S*s were administered, during the same testing session, the Traditional Family Ideology (TFI) scale (Levinson & Huffman, 1955), the F(ascism) scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), and the Value Achievement (vAch) scale (deCharms, Morrison, Reitman, & McClelland, 1955).

The TFI scale is an index of authoritarian ideology within the framework of family ideals. It is a Likert-type scale consisting of 40 items, and it is highly correlated with the F(ascism) scale (see Becker. 1958). In the present study the TFI was buffered with seven items modified from the MMPI *L* scale.

The vAch scale is also a Likert-type scale, composed of nine items concerning attitudes toward achievement. The *S*s who score high on the vAch scale have been characterized as being responsive to authoritative opinions, and as showing explicit motives to achieve for external rewards rather than for the internal satisfaction of achievement (deCharms et al., 1955).

---

[2] It should be noted that in order to obtain the present sample of patients meeting both clinical and psychometric criteria, more than 100 patients who satisfied the clinical criteria had to be rejected for failing to meet the psychometric criteria.

On each of the Likert-type scales (TFI, vAch, F scale), *S*s indicated their response to each item on a 6-point continuum ranging from "extreme agreement" to "extreme disagreement."

Conformity measure. Two days after administering the test battery, *E* returned to the hospital in order to interview *S*s individually. At this time each *S* was administered a 15-item inventory, constructed of 8 TFI items and the previously mentioned 7 buffer items modified from the MMPI *L* scale. These 15 items were selected because pilot research indicated that they received approximately equal agreement-disagreement scale values from nonschizophrenic depressed patients and nondepressed schizophrenic patients.

The *E* had a tape recording with her, and just prior to giving the 15-item inventory to an *S* she said in a conversational tone:

> I would like you to listen to this tape recording with me. You will hear me read each item and then you will hear three women give their opinions to each item according to the scale in front of you. These women are very much like you. They feel as you do and have been in this hospital. I would like you to listen to each item and the opinions expressed and then I want you to tell me your opinion for each item.

The recording contained *E* reading each of the 15 items and three confederates[3] responding to them. On five TFI items and five *L* scale items the confederates commented freely and gave responses which averaged out to the center of the agreement-disagreement scale. The more extreme opinions they expressed on the three remaining TFI items and the two remaining *L* scale items are indicated in Table 1. Although the confederates were "programmed" to make specified responses on the items, their commentary was ad libbed.[4]

After each item and its three responses were heard, the recording was stopped and *S* was instructed to respond verbally in some detail, and then to check off a point on the scale containing the printed item. In this manner it was possible to compare each *S*'s score on these 15 items before and after hearing the tape recording.

TABLE 1

15-ITEM INVENTORY USED IN THE CONFORMITY SITUATION

1. A person's table manners should be just as good at home as when she is out in company.[a]
2. The family is a sacred institution, divinely ordained.
3. Women should take an active interest in politics and community problems as well as in their families.
4. It is only natural and right for each person to think that his family is better than any other.[a]
5. Women think less clearly than men and are more emotional.[a]
6. One should never think of things too bad to talk about.
7. Never put off until tomorrow what you should do today.
8. A person should never gossip.
9. A wife does better to vote the way her husband does, because he probably knows more about such things.
10. One should like everyone she knows.[b]
11. One should never get angry.
12. Women can be too bright for their own good.
13. A person should never vote for men about whom she knows little.
14. One of the worst problems in our society today is "free love" because it mars the true value of sex relations.
15. The unmarried mother is morally a greater failure than the unmarried father.[b]

[a] Item answered in extreme disagreement.
[b] Item answered in extreme agreement.

RESULTS

The results will be presented in two parts. First the performance of *S*s in the D and Sx groups in the conformity situation will be compared. Then the responses of the two groups to the various self-report measures will be evaluated.

Conformity Situation

First, the pretest responses to the five critical conformity items were analyzed in order to evaluate the possibility that the D and Sx groups differed in response tendency at the outset of the experiment. Each *S*'s response to each of the five critical items was converted to a positive number ranging from 1 to 6, depending on the degree of agreement. "Extreme agreement" was assigned a scale value of 6, and "extreme disagreement" a value of 1. The mean score for the five items combined for *S*s in the D group was 4.00 with a standard deviation of 1.87; the mean for *S*s in the Sx group was 4.27 with a standard deviation of 1.78. When these means were compared no difference was found between them ($t = 0.15$, $df = 19$). Furthermore, no differences were found between groups in responses to any of the items individually. Thus, the observation that there was no dif-

---

[3] The authors are indebted to Arline Fisher, Leonore Ganschow, and Jean Malaney for serving as confederates.

[4] Confederates were asked to ad lib rather than read from a script, so that their responses would sound fluent and natural.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE 2 GROUPS ON THE 4 SELF-REPORT MEASURES

Measure	Depressed group		Schizophrenic group		Test of significance
	M	*SD*	*M*	*SD*	
vAch	37.6	9.03	39.0	7.80	$t = 0.38$
F scale	120.4	14.07	123.5	34.10	$U = 52$
TFI	158.3	27.27	157.4	26.47	$t = 0.08$
ESD	24.2	0.79	32.5	2.30	$U = 4$*

* $p < .001$.

ference between groups at the outset of the experiment was confirmed.

Difference scores between the pretest and posttest responses to the five critical items were then compared for *S*s in the D and Sx groups. These scores were derived in the following manner: if an *S*'s response in the posttest situation was in the direction of the responses heard on the tape recording, the numerical difference between her posttest and pretest responses was scored plus; if the response was in the opposite direction, the difference was scored minus; and if *S*'s response did not change from pretest to posttest, the difference was assigned a zero score. The mean difference score on the five critical items for *S*s in the D group was found to be +1.80 ($SD = 0.97$), while the mean difference score for *S*s in the Sx group was +0.25 ($SD = 0.64$). The difference between these means was found to be statistically significant ($t = 4.34$, $df = 19$, $p < .01$), confirming the first hypothesis that depressed *S*s would show more conformity than psychiatric control *S*s.

Response set and response variability. No hypotheses were formulated about response shifts from pretest to posttest on the 10 noncritical (unbiased) items. It was possible, however, to explore the extent to which *S*s in the two groups manifested agreement-disagreement response sets and general response variability to these items.

In order to evaluate relative tendencies to agree or disagree after hearing the tape, difference scores for the 10 noncritical items were evaluated in the following manner: a numerical difference between a posttest response and a pretest response which tended toward greater *agreement with the item* in question was scored plus, while a shift which tended toward *disagreement with the item* was scored minus. Using this method it was found that the mean relative shift for *S*s in the D group was −0.18 ($SD = 0.78$), and the mean for *S*s in the Sx group was −0.10 ($SD = 0.57$). No difference between these means was found ($t = 0.10$, $df = 19$), nor were either of these means significantly different from zero. Thus, it was concluded that neither group showed any tendency to shift consistently in an agreeing or disagreeing manner on the noncritical items.

In order to evaluate the extent to which *S*s showed absolute variability in the shift from pretest to posttest, irrespective of agreement or disagreement, any shift in response, regardless of direction, was scored positively. The mean absolute shift for *S*s in the D group was found to be 1.56 ($SD = 0.70$), and the mean absolute shift for *S*s in the Sx group was found to be 0.89 ($SD = 0.34$). The difference between these means was statistically significant ($t = 2.73$, $df = 19$, $p < .02$), indicating that *S*s in the D group were more variable on items for which no social consensus had been presented.

Self-Report Measures

The means and standard deviations on the four self-report measures for *S*s in the D and Sx groups are presented in Table 2. The differences between groups on the vAch scale and the TFI scale were evaluated by *t* tests, but since the assumption of homogeneity of variance was not met for the distribution of difference scores on the F(ascism) scale, and the ESD scale, these were evaluated by Mann-Whitney *U* tests. As may be noted in Table 2, a significant difference between groups was obtained only for the ESD scale, in which *S*s in the Sx group obtained higher mean scores than *S*s in the D group, contrary to expectation. Thus, the Subsidiary Hypotheses *a*, *b*, and *c* were not supported, and the data bearing on Subsidiary Hypothesis *d* suggest that *S*s in the D group, contrary to expectation, subscribe less than control *S*s to responses which reflect social desirability.

DISCUSSION

The depressive and manic-depressive person has been described as one who has an excessive tendency to conform to social expectation. However, his conformity behavior usually has been assessed indirectly by inference from self-reports presumed to be related to conformity tendencies (Becker, 1960; Cohen et al., 1954; Spielberger et al., 1963). It is not known how well scores on such tests (e.g., TFI, F(ascism) scale) actually predict conformity behavior such as opinion change in the face of social pressure. In the present study, therefore, conformity was assessed in a situation adapted from Asch's (1958) work on the effect of group pressure on the modification of judgments. The present study found that when such a technique was employed, depressed *S*s showed a greater tendency to conform to social pressure than did a matched group of schizophrenics. However, when conformity and achievement-related characteristics of these patients were assessed by the use of self-report inventories, the results were either nonsignificant or opposite to those found in the conformity situation.

Conformity Behavior

The present finding that patients in the D group manifested more conformity behavior than control *S*s is consistent with expectations based upon Becker's (1960) findings with manic-depressed patients; however, the additional finding that these patients also manifested greater response variability on items where no conformity pressure was evident has not been reported before. These two findings, taken together, may be interpreted in two ways. First, it is possible that depressed patients characteristically shift their verbalized attitudes and opinions in a trial-and-error fashion until reinforcement is obtained; and further, that verbalizing attitudes which conform with those of others is secondarily reinforcing, since perceived conformity may have been associated previously with security and acceptance. A second interpretation may be formulated with respect to the characteristics of the control group rather than the D group. The *S*s in the Sx group showed virtually no response shifting, in either an absolute or directional sense. It is possible, therefore, that the significant difference between these two groups may have been determined either by abnormally low variability in the Sx group, high variability of the D group, or both. These alternatives could be studied in future work by comparing both D and Sx groups with a nonpsychiatric control group.

Self-Report Measures

vAch, F(ascism), and TFI scales. The present study failed to support the hypotheses that depressed patients would obtain higher scores than controls on the vAch scale, the F(ascism) scale, and the TFI scale. The failures on the vAch scale and the F(ascism) scale are consistent with recent findings obtained with nonpsychiatric cyclothymic *S*s (Becker & Nichols, 1964), while the failure on the TFI scale is consistent neither with previous findings for manic-depressives (Becker, 1960) nor cyclothymics (Becker & Nichols, 1964). These findings, when taken together are somewhat confusing, and it appears that for the population sampled in the present study no conclusions concerning value achievement, authoritarianism, and traditional family ideology as assessed by these scales can be drawn safely.

ESD scale. The present finding that *S*s in the D group obtained lower scores on the ESD scale than *S*s in the Sx group, while unexpected, is consistent with recent interpretations of the ESD scale (as well as other SD scales) as a measure of *defensiveness* rather than "social desirability" (Ford, 1964a, 1964b). On the assumption, then, that the ESD is an index of defensiveness, it might be expected that depressed patients, who are presumably experiencing pronounced feelings of guilt and self-devaluation, would be likely to be self-derogating and nondefensive in their approach to self-report inventories, thus scoring low on the ESD. This interpretation derives support from Edwards' (1965) recent report of a negative correlation (−.65) between the MMPI *D* scale and the ESD scale.

The difference between ESD scores for *S*s in the D and Sx groups may also be a function of *high* defensiveness of *S*s in the Sx group, most of whom were diagnosed as "paranoid-type." It is quite consistent with the behavior of paranoid patients to be extremely defensive in their self-reports and especially in their approach to personality tests of the MMPI variety. Thus, it would not be unusual for such patients to receive rather elevated scores on a scale which reflects defensiveness in self-description.

In summary, while the present investigation failed to confirm that depressed patients share with manic-depressives the characteristics of high vAch, authoritarianism, or excessive adherence to traditional family ideology, it demonstrated that depressed patients share with manic-depressives a propensity for conforming behavior. Further, rather than assessing conformity by inference from self-report inventories this study assessed it in a situation where *S*s were exposed to group pressure to conform.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ARIETI, S. Manic-depressive psychosis. In S. Arieti (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1959. Pp. 419–454.

ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (3rd ed.) New York: Holt, 1958. Pp. 174–183.

BECKER, J. *An analysis of manic-depressive character in terms of achievement and conformity*. (Doctoral dissertation, Duke University) Ann Arbor, Mich.: University Microfilms, 1958, No. 58-3389.

BECKER, J. Achievement related characteristics of manic-depressives. *Journal of Abnormal and Social Psychology*, 1960, **60**, 334–339.

BECKER, J., & NICHOLS, C. H. Communality of manic-depression and "mild" cyclothymic characteristics. *Journal of Abnormal and Social Psychology*, 1964, **69**, 531–538.

BLUM, D. L. *A behavioral analysis of conformity-related characteristics of depressed patients*. Unpublished master's thesis, State University of New York at Buffalo, 1966.

COHEN, M. B., BAKER, G., COHEN, R. A., FROMM-REICHMANN, F., & WEIGERT, E. V. An intensive study of twelve cases of manic-depressive psychosis. *Psychiatry*, 1954, **17**, 103–137.

DECHARMS, R., MORRISON, W., REITMAN, W., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 414–423.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

EDWARDS, A. L. Correlation of "a unidimensional depression scale for the MMPI" with the *SD* scale. *Journal of Consulting Psychology*, 1965, **29**, 271–273.

FINLEY, C. B., & WILSON, C. D. The relation of the family to manic-depressive psychosis. *Diseases of the Nervous System*, 1951, **12**, 39–43.

FORD, L. H., JR. A forced-choice, acquiescence-free, social desirability (defensiveness) scale. *Journal of Consulting Psychology*, 1964, **28**, 475. (a)

FORD, L. H., JR. Social desirability, defensive denial, and expectancy for success. Paper read at Midwestern Psychological Association, St. Louis, May 1964. (b)

GIBSON, R. W., COHEN, M. B., & COHEN, R. A. On the dynamics of the manic-depressive personality. *American Journal of Psychiatry*, 1959, **115**, 1101-1107.

LEVINSON, D. J., & HUFFMAN, P. E. Traditional family ideology and its relation to personality. *Journal of Personality*, 1955, **23**, 251–273.

MALZBERG, B. Important statistical data about mental illness. In S. Arieti (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1959. Pp. 161–174.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

SPIELBERGER, C. D., PARKER, J. B., & BECKER, J. Conformity and achievement in remitted manic-depressive patients. *Journal of Nervous and Mental Disease*, 1963, **137**, 162–172.

(Received August 12, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 413–415

OPERANT EYELID CONDITIONING IN TRISOMY-18

E. DOROTHY BROWNFIELD AND JACK D. KEEHN
Sonoma State Hospital *Washington State University*

2 profoundly retarded children with extra No. 18 chromosomes were conditioned to emit operant eyeblinks in conformity with continuous and extinction schedules of reinforcement. Conditioning with each child was undertaken at normal feeding times, and when spoonfuls of food were made contingent upon blinking, blink rates increased. On interpolated and final sessions when food was given independently of blinking, blink rates fell to baseline level. This is the 1st reported evidence of learning in the trisomy-18 condition, probably because most of these children die before 1 yr. of age.

Reports of cases now taken to be autosomal trisomy-18 were first published in 1960 by Edwards, Harnden, Cameron, Crosse, and Wolff (1960), Patau, Smith, Therman, Inhorn, and Wagner (1960), and by Smith, Patau, Therman, and Inhorn (1960). Since then several other cases have been reported and the overall physical anomalies have been summarized by Smith, Patau, Therman, and Inhorn (1962) and Hecht, Bryant, Motulsky, and Giblett (1963). Other chromosome 18 aberrations have been reviewed by Grouchy (1965). An inevitable accompaniment of trisomy-18 is "apparent mental retardation," but formal psychological data have not been reported, probably because nearly all cases have died before the age of 1 year, and because the babies are almost completely unresponsive.

To study learning in cases of trisomy-18 the operant eyelid-conditioning technique described by Keehn, Lloyd, Hibbs, and Johnson (1965) was employed. With this technique reinforcements are delivered contingent upon some predetermined pattern of eyeblinking and cumulative blink records compared with baseline data. Keehn et al. (1965) have reported successful conditioning of eyeblinking without awareness in college students. The procedure is particularly apposite for the present study, for the subjects (*S*s) had no visual motor control, but did blink.

METHOD

Subjects. The two cases, both female, appear to have thrived better than is normal for trisomy-18 babies. One, *Tammy*, was 4 years, 11 months at the beginning of the study, and the other, *Tootles*, was 2 years, 8 months. Neither child had previously shown any responsiveness or evidence of learning, and each was reported as profoundly retarded with estimated IQs of 10. Both children lay passively in their cribs except when handled for feeding, bathing, or other routine matters, and could reasonably be described as vegetative (Fuller, 1949). Tammy weighed 19 pounds, 9 ounces and was 34 inches long; Tootles weighed 14 pounds, 11 ounces and was 31 inches long. All these measures are well below the third percentile for normal babies of the ages of these *S*s (Nelson, 1964). In fact, Tootles and Tammy, respectively, weighed close to the third and fiftieth percentiles for 9-month-old infants.

Procedure. The *S*s were normally fed three times daily, at about 7:30 A.M., 12 noon, and 4:30 P.M. Conditioning was attempted at the midday and afternoon feedings, with food as reinforcement. Under regular ward routine *S*s were spoon fed pureed food while held either sitting or lying down, and then given a supplementary bottle feeding in their cribs. During the experiment one experimenter (*E*) took over the spoon feeding of the infants.

In two preexperimental sessions the rate of eyeblinking of each *S* while being fed normally was recorded. Then baseline blink rates were measured over two experimental sessions with *E* feeding *S*s. In these sessions feeding was not contingent upon blinking. For the next 8 or 10 sessions feeding was contingent upon eyeblinking (there was one interpolated extinction session for Tootles and once she slept through the session). That is, *E* held the loaded spoon before *S* and presented it whenever *S* blinked. Technically this is a continuous reinforcement (or FR1) schedule, but occasionally *S* would blink too fast for every blink to be reinforced, and sometimes the food was refused. For 3 more days the *S*s were fed by the usual ward personnel without regard for eyeblinking, and finally both *S*s were given four more experimental sessions in the order: extinction, conditioning, conditioning, extinction. Tammy received two additional extinction sessions.

Before each session the meal was tested for palatability, and whenever *S* refused to eat the diet was modified, usually by adding some pureed fruit. When *S* accepted the food the session was begun. Each

TABLE 1

SESSION BLINKS PER MINUTE OF 2 TRISOMY-18 INFANTS UNDER THE DESIGNATED EXPERIMENTAL CONDITIONS

Subject	Condition	Tammy	Tootles
Age		4.11	2.8
IQ (Estimated)		10	10
Session			
—	Preexperiment	1.6	—
—	Preexperiment	3.3	1.9
1	Baseline	3.0	1.1
2	Baseline	1.5	1.5
3	FR1	3.7	0.8
4	FR1	1.6	1.3
5	FR1	2.7	2.8
6	FR1	3.7	2.7
7	FR1	3.7	4.2
8	FR1	3.8	3.1
9	FR1	3.2	3.2
10	FR1	6.7	—
	Extinction	—	1.7
11	FR1	5.4	3.0
12	FR1	4.7	—
13	Extinction	2.6	1.5
14	FR1	5.0	3.0
15	FR1	4.6	2.6
16	Extinction	4.1	1.5
17	Extinction	4.4 (2.8)	—
18	Extinction	3.0 (2.7)	—

Note.—On Session 17, 25 of Tammy's 66 blinks were squints, and on Session 18, 5 of her 45 blinks were squints. The figures in parentheses are blink rates with squints omitted.

session lasted 15 minutes. On extinction sessions food was not withheld, but, rather, was not contingent upon blinking. The times at which *S* blinked were recorded against a running stopwatch.

RESULTS

Table 1 shows the rate of blinking of each *S* on the preexperimental trials when *E* observed *S*s while they were being fed normally by an attendant, on the two baseline sessions when one *E* fed *S*s independently of blinking, on the conditioning trials when feeding was contingent upon blinking, and on the extinction trials when feeding was again independent of blinking. Odd-numbered sessions are noontime feedings; even sessions are corresponding afternoon feedings. There is no record for Tootles on Session 12 as she was too sleepy to feed at the time the session was scheduled.

On the preexperimental observation Tootles' blink rate was 1.9 blinks per minute. This low rate was maintained in Baseline Sessions 1 and 2, when blink rates were 1.1 and 1.5 per minute, respectively, and in Conditioning Sessions 3 and 4 when rates were 0.8 and 1.3 per minute, respectively. However, over the next 5 conditioning sessions (Sessions 5 through 9) blink rates were 2.8, 2.7, 4.2, 3.1, and 3.2 responses per minute. On Session 10, an interpolated extinction session, the blink rate fell to the baseline level of 1.7 blinks per minute, only to rise to 3.0 blinks per minute when blinking was again reinforced on Session 11. On the final 4 sessions, 13 through 16, when extinction and conditioning were counterbalanced, blink rates on successive sessions were 1.5 (extinction), 3.0 (conditioning), 2.6 (conditioning), and 1.5 (extinction) blinks per minute.

Tammy's preexperimental blink rates were 1.6 and 3.3 blinks per minute on two successive observations. These rates were maintained at 3.0 and 1.5 blinks per minute in Baseline Sessions 1 and 2. Rates of about 3.7 blinks per minute or less were typical on Sessions 3 through 9. On Sessions 10, 11, and 12, blink rates were 6.7, 5.4, and 4.7 blinks per minute, and on the 4 sessions, 13 through 16, when extinction and conditioning were counterbalanced, blink rates on successive sessions were 2.6 (extinction), 5.0 (conditioning), 4.6 (conditioning), and 4.1 (extinction) blinks per minute. Because the final extinction rate was above baseline level two further extinction sessions were run, in which blink rates of 4.4 and 3.0 blinks per minute were recorded. This last rate is within the range of baseline rates recorded at the beginning of the experiment. On both of the final sessions several of Tammy's blinks were squints in which she screwed up her whole face. If these squints are not counted as blinks then blink rates on these extinction sessions were 2.8 and 2.7 blinks per minute, respectively.

DISCUSSION

There is little doubt that the reinforcement-by-feeding contingency did modify the eyeblinking rates of these infants: when feeding immediately followed blinking the blink rates of both *S*s rose, and when the contingency was removed blink rates fell. These findings add to those of Fuller (1949) in demonstrating operant conditioning of vegetative human *Ss*, and to the rapidly expanding literature on the clinical use of operant conditioning principles. They also demonstrate that when infants with the trisomy-18 condition survive beyond their first year they are capable of learning, and perhaps are not so profoundly and permanently retarded as formerly reported to be.

REFERENCES

EDWARDS, J. H., HARNDEN, D. G., CAMERON, A. H., CROSSE, V. M., & WOLFF, O. H. A new trisomic syndrome. *Lancet,* 1960, **1**, 787–790.

FULLER, P. R. Operant conditioning of a vegetative human organism. *American Journal of Psychology,* 1949, **62**, 587–590.

GROUCHY, J. DE. Chromosome 18: A topologic approach. *Journal of Pediatrics,* 1965, **66**, 414–431.

HECHT, F., BRYANT, J. A., MOTULSKY, A. G., & GIBLETT, E. R. The No. 17-18 (E) trisomy syndrome. Studies on cytogenetics, dermatoglyphics, paternal age and linkage. *Journal of Pediatrics,* 1963, **63**, 605–621.

KEEHN, J. D., LLOYD, K. E., HIBBS, M., & JOHNSON, D. Operant eyeblink conditioning without awareness: A preliminary report. *Psychonomic Science,* 1965, **2**, 357–358.

NELSON, W. E. (Ed.) *Textbook of pediatrics.* (8th ed.) Philadelphia: Saunders, 1964.

PATAU, K., SMITH, D. W., THERMAN, E., INHORN, S. L., & WAGNER, H. P. Multiple congenital anomaly caused by an extra autosome. *Lancet,* 1960, **1**, 790–793.

SMITH, D. W., PATAU, K., THERMAN, E., & INHORN, S. L. A new autosomal trisomy syndrome. *Journal of Pediatrics,* 1960, **57**, 338–345.

SMITH, D. W., PATAU, K., THERMAN, E., & INHORN, S. L. The No. 18 trisomy syndrome. *Journal of Pediatrics,* 1962, **60**, 513–527.

(Received August 20, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 416

SIZE OF FIGURE DRAWINGS OF DEPRESSED PSYCHIATRIC PATIENTS

HOWARD B. ROBACK[1] AND ALFRED L. WEBERSINN
Hawthornden State Hospital, Northfield, Ohio

2 studies testing the hypothesis that depressed persons draw small, restricted figure drawings were carried out. Their findings revealed an inconsistent relationship between depression and size of figure drawings.

Lewinsohn's (1964) research empirically supports the impression of Rapaport, Gill, and Schafer (1946) and Machover (1949) that persons in a depressed state are likely to draw human figures that are smaller than those of persons not in such a state. This difference has been attributed to the reduced productivity and minimal energy expenditure that characterizes depressed persons. The present study was a further attempt to experimentally compare the size of figure drawings of depressed and nondepressed psychiatric patients.

The current investigators carried out two studies utilizing independent criteria of depression. In Study 1 the height of same-sex drawings of a random selection of 25 male and 25 female patients with *T* scores above 67 on the *D* scale of the MMPI (regardless of other scores) was compared with same-sex drawings of a random selection of 25 male and 25 female patients whose *T* score on the *D* scale of the MMPI was less than 67. In Study 2 the height of same-sex drawings by patients who were considered depressed by their doctors at the time of their admission to the hospital was compared with figure drawings of patients not considered in this state—a procedure used by Lewinsohn (1964).

The results of both studies, which are presented in Table 1, reveal that the depressed groups' drawings generally tend to be smaller than those of the nondepressed group; however, these results were not statistically significant. In Study 1 there was a significant difference ($p < .01$) in the height of drawings between depressed and nondepressed female patients, the depressed females having the smaller drawings. However, similar findings were not found between drawings of depressed and nondepressed males in Study 1, or between any groups in Study 2.

TABLE 1

HEIGHT IN MILLIMETERS OF FIGURE DRAWINGS OF DEPRESSED AND NONDEPRESSED GROUPS

	Depressed group			Nondepressed group			
	N	*M*	*SD*	*N*	*M*	*SD*	*t*
Study 1	50	136.44	58.52	50	152.80	50.89	*ns*
Study 2	46	123.39	67.45	49	132.57	55.24	*ns*

The findings of this study, which reveal an inconsistent relationship between depression and size of figure drawings, do not support those of Lewinsohn (1964). However, the significant difference in height of same-sex drawings between depressed and nondepressed female patients used in Study 1 ($p < .01$), and the overall trends in both studies, suggest that more research with the Draw-A-Person test appear in order. It may be that variables other than depression result in small figure drawings. For example, Gray and Pepitone (1964) found that subjects with low self-esteem draw small figures although they are not depressed.

REFERENCES

LEWINSOHN, P. Relationship between height of figure drawings and depression in psychiatric patients. *Journal of Consulting Psychology*, 1964, **28**, 380–381.

MACHOVER, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.

PEPITONE, A., & GRAY, D. M. Effect of self-esteem on drawings of the human figure. *Journal of Consulting Psychology*, 1964, **28**, 452–455.

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Chicago, Ill.: Yearbook Publication, 1946.

(Received September 10, 1965)

---

[1] Now at York University. The authors are indebted to Regina Hunter and Bonny Ward for their assistance in the collection of the data and to V. Bixenstine, Kent State University, for reading this manuscript.

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 417–419

COPING AND AVOIDANCE IN INEBRIATED ALCOHOLICS AND NORMALS[1]

DONALD POLLACK

University of California, Los Angeles

Inebriation significantly decreased the coping behavior of alcoholics as measured by a Sentence Completion Test. Normals showed the opposite pattern of change. It was inferred that alcohol therefore acts to release behaviors which are latent and interdicted in the sober state for both groups. Suggestions for future research are made.

A methodological advance in studying a given abnormal behavior would be the production, under controlled observation, of the deviant behavior itself. Work with alcoholics who are inebriated provides such an opportunity for study.

The literature on experimental intoxication of subjects (*S*s) is rather sparse (Gantt, 1957; Hartocollis, 1962; Kawi, 1961; Newman, 1935; Takala, Pikkanen, & Markkanen, 1957), and alcoholics have not been used as *S*s. Therefore it was considered desirable to experimentally intoxicate alcoholics and normals for the purposes of measuring and comparing behavioral alterations. For this purpose a revised Sentence Completion Test (SCT) initially devised by Mainord (1956) was utilized, which is scorable on a coping-avoidance dimension. The SCT attempts to separate individuals who (*a*) try to cope with emotionally charged stimuli, from (*b*) individuals who characteristically avoid or evade those situations. Thus coping behavior involves maturity, independence, and commitment to the demands of reality. Regression, passivity, and withdrawal signal a failure in coping.

The working hypothesis was derived from a role-play study by MacAndrew (1962) which suggested that alcohol intake had the following effects on alcoholics and normals: (*a*) that alcoholics would cope less effectively when inebriated than when sober and, (*b*) that the opposite would hold true for normals.

METHOD

There were 20 alcoholics and 20 normals who constituted the samples. The alcoholics were hired from a State Employment Agency and had a history of economic, social, and physical losses due to drinking. The normal control group was recruited mainly from the custodial staff of a local university, and was matched for age, sex, education, and race.

The alcoholics were given 1.25 cubic centimeters per kilogram of body weight of 190-proof alcohol, administered orally in an orange juice solution. After blood alcohol levels of between .10 and .15% were reached, the SCT was administered. Since the normals could not function effectively above the .05% level, this level was maintained for them. The Breathalyzer[2] was used to measure these levels. SCT testing was counterbalanced, in that, on the same day, half of each group was initially tested sober, then while inebriated. A Lindquist Type VI repeated-measures design was utilized for the analysis of variance (Lindquist, 1956).

The specially developed form of the SCT contains 60 items, 40 of which are scorable with 20 filler items. Goldstein (1957) modified the SCT slightly and increased its reliability by the development of a scoring manual. Scoring is on a 3-point scale, 0, 1, 2, on a continuum from avoidance (0) to direct and personal responding (2). The following rules are used to determine scoring:

1. The more personal and specific the response, the higher the score.
2. The less arbitrary or stereotyped the response, the higher the score.
3. The stronger the feeling expressed, the higher the score.
4. Items left blank, or answered "I don't know," are automatically scored 0.

Scorer and test reliability were both in the .80s for this study. An *S*'s score is then the arithmetic sum of his individual item scores. Item content is weighted strongly in the direction of sexual and aggressive themes, thus adding to the impact of admittedly projective stimuli.

---

[1] This paper represents a principal part of a dissertation submitted in partial fulfillment for the doctoral degree at the University of California, Los Angeles. Special thanks are due to George F. J. Lehner for his guidance and encouragement.

[2] The Breathalyzer was developed by Borkenstein (1959) and utilizes potassium dichromate as the reagent for oxidizing alcohol in alveolar air.

RESULTS

The SCT data shown in Table 1 are statistically analyzed in Table 2. As expected, the significant results (at the .001 level) appear in the interaction of nosology and drunk-sober conditions. Thus the differing pattern of reactivity to alcohol as shown by each group becomes central, with an examination of these specific changes, when appropriate, via *t* tests. However, the overall conclusion is that the interaction very clearly shows that alcoholics "cope" more when sober than they do when drunk. Normals show the opposite pattern of change, as their "coping" scores increase under the alcohol condition. The main effects of nosology and inebriation are therefore not significant because of the interactive effect. Subsequent *t* tests also indicated that the alcoholics coped more than normals when both were sober ($t = 7.34$), but coped less when both were drunk ($t = 10.19$).

The order of position on the continuum from highest "coping" to most "avoidance" was (*a*) normals drunk, (*b*) alcoholics sober, (*c*) normals sober, and (*d*) alcoholics drunk.

TABLE 1

SENTENCE-COMPLETION SCORES FOR ALCOHOLICS AND NORMALS (SOBER AND DRUNK)

Subject	Alcoholics		Normals	
	Sober	Drunk	Sober	Drunk
1	37	30	35	38
2	49	37	33	38
3	31	17	31	41
4	36	23	32	41
5	40	25	32	46
6	41	36	44	52
7	39	35	27	37
8	36	33	26	33
9	36	32	22	36
10	36	26	23	35
11	39	38	37	42
12	45	33	38	42
13	40	35	31	40
14	43	33	30	44
15	48	39	27	35
16	38	27	29	37
17	44	39	36	45
18	24	28	37	46
19	42	36	41	48
20	41	33	41	46
Mean	39.3	31.8	32.6	41.1

TABLE 2

SENTENCE-COMPLETION ANALYSIS OF VARIANCE

Source of variation	*SS*	*df*	*MS*	*F*
Between subjects		39		
Alcoholic–Normal	36.45	1	36.45	.70
Order	245.00	1	245.00	4.73*
Alcoholic–Normal × Order	0.00	1	0.00	0.00
Error	1866.10	36	51.84	
Within subjects		40		
Drunk–Sober	5.00	1	5.00	.10
Order × Drunk–Sober	1.25	1	1.25	.15
Alcoholic–Normal × Drunk–Sober	1280.00	1	1280.00	24.69**
Alcoholic–Normal × Order × Drunk–Sober	18.05	1	18.05	2.17
Drunk–Sober × Subjects within groups	299.70	36	8.33	

* $p < .05$.
** $p < .001$.

DISCUSSION

Accounting for the effects of alcohol as measured here raises many problems. It seems that withdrawal from responsibility and commitment is a concomitant of inebriation for the alcoholic. On another level, if the traits of passivity and collapse of coping are assumed to be subclinical or latent, then alcoholic ingestion can be seen as a potent agent for rearranging habit-family hierarchies (Hull, 1943). The explanation that alcohol may increase coping for the alcoholic, but against internal pressures, as contrasted with the normal person facing external pressures in a more adaptive way due to alcohol, is intriguing; it matches the clinical picture of inner pressure and turmoil in the alcoholic, which is not true of the normal.

An immediate research proposal suggested would be an attempt to identify the prealcoholic by use of the experimental procedure here (e.g., Army draftees might be a sample, with a collapse of coping being diagnostic).

One other implication of this study would be the evidence against a so-called alcoholic personality. Alcohol could be used by many different personality types to release many different kinds of behaviors, in other words as a final common pathway. This enabling function of alcohol would allow for behaviors to appear which would otherwise be interdicted in the sober state. Since anxiety reduction via inebriation (as a reinforcement) was not strongly evidenced, this must be considered. The complex nature of initial relief of anxiety and tension due to alcohol, followed by entry into previously avoided situations, probably leads to a secondary increase in anxiety. Thus, that alcohol may provide potent reinforcements by releasing lower probability behaviors seems more rewarding for understanding the behavior of the alcoholic, as compared with a direct anxiety-reduction model (Pollack, 1965).

It seems fair to conclude that the same drug, acting on different personality structures, produces different results.

REFERENCES

BORKENSTEIN, R. F. The breathalyzer—manual of instructions. Red Bank, N. J.: Stephenson Corporation, 1959.

GANTT, W. H. Acute effect of alcohol on autonomic (sexual, secretory, cardiac) and somatic responses. In H. E. Himwich (Ed.), *Alcoholism: Basic aspects and treatment*. (Publ. No. 47) Washington, D. C.: American Association for the Advancement of Science, 1957. Pp. 73–89.

GOLDSTEIN, M. The relationship between coping and avoiding behavior and attitude modification. Unpublished doctoral dissertation, University of Washington, 1957.

HARTOCOLLIS, P. Drunkenness and suggestion: An experiment with intravenous alcohol. *Quarterly Journal of Studies on Alcohol*, 1962, **23**, 376–389.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

KAWI, A. A. Effects of sodium amytal and ethyl alcohol on some psychomotor functions. *Journal of Neuropsychiatry*, 1961, **2**, 195–200.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.

MACANDREW, C., & GARFINKEL, H. A consideration of changes attributed to intoxication as common-sense reasons for getting drunk. *Quarterly Journal of Studies on Alcohol*, 1962, **23**, 252–266.

MAINORD, W. A. Experimental repression related to coping and avoidance in the recall and relearning of nonsense syllables. Unpublished doctoral dissertation, University of Washington, 1956.

NEWMAN, H. W. Alcohol injected intravenously: Some psychological and psychopathological effects in man. *American Journal of Psychiatry*, 1935, **91**, 1343–1352.

POLLACK, D. Experimental intoxication of alcoholics and normals: Some psychological changes. Unpublished doctoral dissertation, University of California, Los Angeles, 1965.

TAKALA, M., PIKKANEN, T. A., & MARKKANEN, T. *The effects of distilled and brewed beverages*. Helsinki: Finnish Foundation for Alcohol Studies, No. 4, 1957.

(Received September 20, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 420

STABILITY OF FIELD DEPENDENCE IN CHRONIC ALCOHOLIC PATIENTS

GERALD GOLDSTEIN AND JOHN W. CHOTLOS

Veterans Administration Hospital, Topeka, Kansas

62 alcoholic patients were administered the Witkin Rod and Frame Test before and after 8–10 wk. of psychiatric treatment. There was significantly greater field independent performance following treatment. These results tend to cast doubt on the hypothesis that performance on the Rod and Frame Test is stable and unaffected by changes in body state.

In a recent article, Karp, Witkin, and Goodenough (1965) have suggested that level of field dependence is relatively stable in alcoholics and is not markedly affected by the subject's being in an intoxicated state. This finding was confirmed both for the Witkin Rod and Frame (RFT) and Body Attitude Tests (BAT), but not for the Embedded Figures Test (EFT). If the Karp, Witkin, and Goodenough hypothesis is correct, then it would be reasonable to assume that alcoholics would not become substantially more field independent on the RFT on retesting following treatment.

At the Topeka Veterans Administration Hospital chronic alcoholic patients are given a 3-month treatment program and are routinely evaluated before and after treatment. As part of their evaluation they receive standard administrations of the RFT. Administration of this test pre- and posttreatment (about an 8–10 week interval) would therefore provide a test of the stability hypothesis.

METHOD

The subjects were 62 male alcoholics, all of whom were patients in the Topeka Veterans Administration Hospital Alcoholism Treatment Program. Pretesting was generally done during the admission week, and posttesting a week prior to discharge. Subjects ranged in age from 29 to 57 years, with a mean age of 43.8 years. Each subject received the RFT using the same procedures as those described by Karp et al. (1965).

TABLE 1

PRE- AND POSTTREATMENT PERFORMANCE OF 62 ALCOHOLICS ON THE RFT

	RFT-R	RFT-L	RFT-U
D^a	−3.93	−5.60	−3.34
t	2.63	4.65	4.06
p	<.01	<.01	<.01

[a] D is equal to the postscore minus the prescore.

RESULTS AND DISCUSSION

The t tests for correlated samples were carried out for each of the test conditions (chair tilted right, left, and upright). Results are presented in Table 1. It can be seen that there was significantly less deviation from the vertical on posttesting than there was on pretesting, in all three conditions. Sign tests (Siegel, 1956) were also performed, with z being equal to 3.17 ($p < .01$) in the upright condition, 2.92 ($p < .01$) in the tilted right condition, and 5.20 ($p < .01$) in the tilted left condition. Therefore, a significantly greater number of subjects changed in the direction of greater field independence than could be expected on the basis of chance. These results tend to cast doubt on the stability hypothesis as it applies to the RFT. It seems likely that performance on the test improves after a period of abstinence from alcohol ingestion.

REFERENCES

KARP, S. A., WITKIN, H. A., & GOODENOUGH, D. R. Alcoholism and psychological differentiation: Effect of alcohol on field dependence. *Journal of Abnormal Psychology*, 1965, **70**, 262–265.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received September 23, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 421–428

ROLE OF STIGMA AND SET IN INTERPERSONAL INTERACTION[1]

AMERIGO FARINA, CHARLES H. HOLLAND, AND KENNETH RING

University of Connecticut

This study focuses on 1 condition which seems to influence the role of stigmata in interpersonal relationships. The results indicate that mental illness and the extent to which a person is held responsible for this stigma determine the amount of pain inflicted upon him and how favorably he is evaluated. Responsibility was varied by having a confederate report to the undergraduate Ss either a typical or an atypical and pathogenic childhood. A bad childhood mitigated the harshness of the treatment accorded the mentally ill person, but for the normal individual this experience appears to be as stigmatizing as mental illness itself. The individual perceived as abnormal either because of mental illness or a poor childhood experience is treated in a harsher manner than the normal. He is also described as less adequate in his performance, although there is no objective basis for this, he is less liked, and Ss prefer no further interaction with him. An interesting implication of comparing the evaluation with the behavior displayed is that people can be induced to behave more favorably while retaining all of their dislike and contempt for the stigmatized person.

When another individual is perceived to be a cripple, mentally ill, or a beggar, many people feel dislike and aversion toward him. However, perhaps because it is learned from early childhood that help and sympathy should be given to those less fortunate, few people feel only repugnance and contempt. At times, one certainly feels sympathy and behaves in a supportive manner. What are the conditions under which stigma tends to elicit sympathy and support as compared to distaste and repugnance? The central purpose of the present study was to investigate one of these conditions. Specifically, favorability of perception and amount of pain inflicted upon an individual as a function of learning that he is normal or mentally ill and believing either that he is personally responsible or that the environment is responsible for his condition was measured. First, an examination was made of the role played by stigmata in interpersonal relationships and then how the set of the individual perceiving the stigma might influence this role was considered.

By stigma is meant any attribute which discredits and lowers the status of an individual once he is known to have this attribute. There are compelling reasons for believing that such conditions play a crucial role in human interaction. Not only do they seem to be fundamentally involved in such large-scale social phenomena as the civil rights movement and national immigration quotas, but their influence is seemingly never absent in the most ordinary of social interactions. As Goffman (1963) has contended, the most fortunate of persons will have his failings since we all fall short of some of society's ideals of personality, appearance, and biographical characteristics. While there undoubtedly are differences among stigmata with regard to their role in interpersonal relationships, they are similar in some important respects as is implied by definition. For example, any stigma would appear to make the typical interaction among people more tense and less pleasant for all involved. These common elements should permit a fair degree of generalization from one stigma to another.

The particular stigma which is the focus of the present study, mental illness, recommends itself for a number of reasons. Attitudes toward the mentally ill have been thoroughly investigated and are strongly negative (Cohen & Struening, 1962, 1964; Ellsworth, 1965; Jones & Kahn, 1964; Manis, Houts, & Blake, 1963; Nunnally, 1961; Struening & Cohen, 1963). Even knowledge of relatively

---

[1] This study was financed by research grants from the National Science Foundation (GS-848) and from the National Institute of Mental Health, United States Public Health Service (M-6167).

minor adjustment problems involving no treatment makes a person less liked than another whose adjustment is more adequate (Jones, Hester, Farina, & Davis, 1959). Moreover, although most studies concerned with mental illness have employed phenomenological measures, studies using behavioral measures have been carried out. These studies (Cohen & Struening, 1964; Farina & Ring, 1965) show that, in addition to expressing generally more negative feelings, people behave differently toward those perceived as mentally ill. However, the interpersonal significance of the behaviors measured in the studies cited is not always easily discernible. Thus, Cohen and Struening (1964) reported that patients housed in hospitals where the employees score high on restrictive attitudes toward mental illness are retained longer than patients in other hospitals. One can plausibly argue that these behaviors stem from benevolent feelings rather than those of dislike and fear. A major purpose of the present study is to obtain a behavioral measure having a rather clear interpersonal meaning. The measure selected was painfulness of shock administered. As Hastorf, Richardson, and Dornbusch (1958) have persuasively argued, studies using behavioral measures are sorely needed if progress is to be made in the area of person perception.

The sets to understand or to judge which were given to subjects (*S*s) are, respectively, quite similar to what Jones and Thibaut (1958) have called the causal-genetic and situation-matching sets. The former set presumably influences a person to try to comprehend another's personality and to view his departures from normative behavior as the result of events not within his control. This is the set which clinical psychologists presumably adopt toward patients whom they are diagnosing. The latter set, on the other hand, leads an *S* to expect certain behaviors as appropriate in a given situation and thus induces him to hold another personally responsible for failures to meet the situation's requirements. It is this kind of set which jury members are believed to take toward a person on trial. It seems reasonable to expect the unfavorable effects of stigma will be mitigated when the stigmatized is not held responsible for his condition as compared to when he is. For the normal person it is unclear what the effects of set might be. Conceivably, when a person achieves a normal status through his own efforts, he might be given greater credit than the individual whose normality was the result of fortuitous circumstances. For this reason, he might be more favorably evaluated in the former than the latter condition. On the other hand, it seems quite possible that normality elicits a quite favorable reaction no matter how achieved and responsibility for his normal status might thus be irrelevant.

METHOD

Subjects. The *S*s of this study were 41 students enrolled in the introductory psychology course at the University of Connecticut. They served as *S*s in order to satisfy a course requirement. Since it has been shown repeatedly that there are marked sex differences in the perception of others (Shrauger & Altrocchi, 1964), only males were used. These *S*s were randomly assigned to one of the four conditions of the study.

Apparatus. The apparatus consisted of a teacher's panel and a learner's panel which were connected by wires and placed in adjoining rooms. The teacher's panel was, ostensibly by chance, always assigned to the naïve *S*. On this panel there was a row of three lights, labeled A, B, and C, which lighted when *S* in the adjoining room (in reality a confederate) pressed the corresponding buttons. The panel also had a light labeled "Finished" by means of which the confederate signaled that he had pressed the buttons the required number of times. This panel was equipped with a button which presumably administered a shock for as long as it was pressed and a dial which permitted the naïve *S* to set the shock intensity at any of 10 levels, from weak to very strong. Finally, the teacher's panel also had a button labeled "Start" which the naïve *S* pressed to begin each trial. The learner's panel permitted the experimenter (*E*) to determine the intensity level of each shock administered and, by means of a timer with which the panel was equipped, to measure the length of each shock in fractions of a second.

Procedure. When a naïve *S* arrived at the experimental setting, the confederate was introduced to him as a student from another introductory psychology section. The confederate was actually one of two drama students with good acting ability and experience.[2] These confederates were both carefully trained to play the same standardized role. After the introduction it was explained to them that as part of the experiment one of them would be required to tell the other something about himself which might be personal in nature and both were asked to agree to keep the information confidential. This was done to make it believable that the confederate would reveal derogatory personal information as he subsequently did for some of the conditions. They were then told the experiment was concerned with interpersonal relations and their role in communicating with another person, and that one would be asked to assume the role of communicator, or teacher, while the other would act as the learner. Both were asked to pick a card out of an envelope to determine who was to be the teacher and who the learner. Since both cards said "teacher," but the confederate routinely claimed to have drawn the learner card, the naïve *S* was always assigned the teacher role.

---

[2] Graduate students in psychology were initially used as confederates. However, they were not able to convince *S*s that they were ex-mental hospital patients and it was found necessary to employ

They were then shown the teacher's panel and told that a similar learner's panel was in the adjoining room. The task, it was explained, required the teacher to communicate two patterns of button presses which he would know but the learner would not know. The first pattern consisted of four button presses, such as C A B A, and the second consisted of five button presses, such as A B B C A. They were told the correct solution required that the buttons be pressed in a particular order and that each button be pressed at least once for each pattern. It was explained that initially the learner must necessarily guess but the teacher could inform him that he was in error by the administration of an electric shock. At this point the confederate protested mildly but agreed to proceed. Attention was called to the shocking apparatus, and the naïve *S* was told that he was free to choose intensities at any of the 10 levels and to administer the shock as long as he wished. The shock at Setting 1 was described as very weak and barely detectable while that at the maximum setting, 10, was described as very strong and extremely painful but not injurious. For each pattern 25 trials were allowed, a trial being defined as the sequence in which the teacher gives the start signal, the learner presses the buttons the required number of times and indicates he is finished, and the teacher then either shocks the learner if he is incorrect or does nothing if he is correct.

In order to introduce the manipulation, it was pointed out to the confederate and *S* that the more a teacher knows about someone, the more effectively and rapidly he can communicate with him. However, the explanation continued, knowledge about some things is more relevant than knowledge about others and for this experiment past results indicated it was important to know two kinds of things about the learner that, initially, might seem surprising. It was important to know, they were assured, about childhood experiences and about current personal adjustment. The confederate was then asked to give both kinds of information to the teacher, particularly information about unique or unusual experiences, and the request was made that he be as candid as possible. The confederate routinely asked for clarification which was given together with reassurances.

The confederate, at this point, described his childhood experiences in one of two ways. For one description (J), he reported them to be normal and typical, thus presumably inducing the naïve *S* to assume a set to judge the confederate. The J description was as follows:

> I guess I had a pretty typical childhood. I went to the grammar school and high school in my neighborhood. What kind of things do you want? . . . . I remember I used to go on summer trips around the country with my parents, places like Grand Canyon or Niagara Falls, like that. They were really exciting. I think I'd like to go back and see those places again some day. I also played Little League baseball for 2 years on a team my father coached. We didn't win too often, but it was really a thrill. I guess that's about it.

For the second description (U), he reported his experiences to have been disturbing and pathogenic, thus presumably leading *S* to try to understand the communicator. The U description was given as follows:

> I would say my childhood was pretty unusual. I used to play a lot of ball . . . but I guess that's not too unusual. Well, my parents got divorced. I was about 11? 12? I don't know. They never seemed to get along. When we would go someplace, like on a picnic, they would get into an argument and we'd wind up coming home early, no one saying a word. When they were divorced I went to live with my grandparents. They were goodhearted but hardly could speak any English. Coming here is the first time I've gotten away from them. That's about it.

As indicated, there were also two descriptions of adult adjustment. For one, the confederate indicated his condition to be normal (N), whereas for the second he revealed that he was stigmatized (S). The N description was:

> Things are going pretty well for me now. My grades are good, and I think I know what I want to do when I finish school. I have trouble getting along with some people but with most I get along O.K., and I have two very close friends. I have been dating quite a bit but I am not too serious with anyone. I would say I am having a pretty good time in college in general, but I also think I'm getting a good education. That's about it, I guess.

The S description was the following:

> As an adult I've had some problems. I haven't found it easy getting along with others and I

---

drama students. The authors would like to express appreciation to Frederick M. Hughes and Stephen B. Gordon who did this work.

> don't really have any close friends. I don't really know if this is what you wanted. . . . Well, I suppose that the most important thing about me is that I've had a kind of nervous breakdown and was in a mental hospital for a while. As far as school is concerned, I like it and I get good grades. Well, I hope this is what you wanted.

Thus, a 2×2 independent groups design was employed with *S*s assigned randomly to each of the four (JN, JS, UN, and US) conditions of the study.

After the information was given, the naïve *S* and the confederate were separated, the latter being taken to the adjoining room where, ostensibly, electrodes were strapped to his hand. The naïve *S* was then given the correct patterns of button presses and told to begin. The same programmed series of presses were given by the confederate to all *S*s in all conditions. For the first series, which required four presses, the confederate "learned" the pattern on Trial 13. However, although he approached the correct solution, he failed to learn the series requiring five presses in the 25 trials allowed. At this point the confederate was returned to the first room and he and the naïve *S* were asked to complete a questionnaire which, they were assured, would be seen only by *E*. The questionnaire was intended to measure the effectiveness of the manipulation as well as to obtain perceptual measures designed to clarify the role of stigma in social relationships. It is described in the Results section.

Following this, all *S*s were interviewed in a standardized manner beginning with very indirect questions and leading to the question, "Is there anything else the experiment might have been after?" If an *S* indicated being suspicious that any aspect of the study was not as it was described to him, he was discarded. Three *S*s suspected the confederate had not told the truth and were dropped for this reason while a fourth was eliminated from the study because he failed to comprehend the instructions. Finally, the deception was revealed, an effort being made to have *S*s understand why it had been necessary to deceive them, and the true nature of the study was carefully explained. It was pointed out to the participating students that while it was impossible to stop them, it would be greatly appreciated if they would not disclose the deception in the study. They were also told that they would be given the results when these became available, a promise subsequently carried out. Of the 41 usable *S*s, 11 were assigned to the JN condition and 10 were assigned to each of the three remaining conditions.

RESULTS

A question which must be answered at the onset is that of the success of the set and stigma manipulations. Of the 16 items composing the postexperimental questionnaire, 6 were concerned with this issue. Of these, 3 were designed as checks on the stigma and the remaining 3 were intended to serve as checks on the set manipulation. With regard to the former, the confederate was described as getting along more poorly with others and as being less well-adjusted when he presented himself as having been mentally ill in comparison to normal ($p < .02$ and $.01$, respectively, according to a chi-square test). A third item, less central to mental illness, asked for an estimation of how well the confederate understood himself and for this item there were no differences among the groups. Checking on the set manipulation was more difficult since the researchers wanted to determine not merely if the childhood experiences described could be accurately reported but whether the groups differed in perceiving the confederate as differentially responsible for his level of adjustment. To this end, one item asked whether the confederate's adjustment was primarily due to environmental experiences or to his own character. Although the differences were in the expected direction (12 *S*s in the U condition as compared to 8 in the J condition choosing environment), they were not statistically significant. A second item asked for an estimation of how much the confederate had profited from his experiences and provided the choices of: a fair amount, very much, and not too much. With regard to this item, it seems plausible to assume that when the environment is perceived as playing a small and insignificant role in an individual's development, *S*s will favor the relatively neutral choice of a fair amount. However, when this role is seen as crucial and central, they will indicate the profit was either great or small. As expected, more *S*s in the J than the U set condition selected the fair amount alternative (15 in comparison to 6, $\chi^2 = 7.4$, $df = 1$, $p < .01$). The third and final item asked if the confederate was himself largely responsible for what he was or not. Consistent with expectations, more *S*s in the J condition judged him responsible than in the U condition (13 to 7). While a chi-square test indicated this difference to be significant at only $p < .10$, a Fisher exact probability test disclosed the probability level to be less than .06. Moreover, the two groups which consideration suggests ought to be most widely divergent, the JN and US groups, differed significantly in the expected direction ($\chi^2 = 4.1$, $df = 1$, $p < .05$). In sum, these results suggest that both stigma and set were successfully manipulated.

Table 1 presents the results of the shock measures. The pattern for the two series is quite consistent, the JN condition being in each case the one where the mildest shock was given and also where it was given for the shortest period of time. The most intense shock was administered in the JS condition. However, it was administered for the longest period in the UN condition. What appears to be a far more intriguing result, however, is the opposite effect of set for a normal as compared to a stigmatized individual. For each measure and in each series a set to understand leads to less pain being inflicted on a stigmatized individual than a set to judge him, whereas the opposite is true for a normal person.

To simplify the analysis of the shock data, one *S* was selected at random from the JN condition and removed, thus leaving 10 *Ss* in each group. The analyses done were as described by Winer (1962, pp. 337–349). This type of analysis is suitable for three-variable experiments like the present one. Each measure for each series was analyzed separately and an analysis was done using a general punishment score obtained by multiplying the intensity by the length of the shock given on each trial. A simple multiplicative score was used in the absence of any a priori reason for combining the two shock measures.

The punishment scores for the first series of button presses did not approach being significantly different as a function of either stigma or set. However, the *F* score for the interaction was 3.26 which, for $df = 1$, indicates a probability level below .10. Comparable results were obtained for the scores of the second series but for this series the interaction *F* reached a statistically significant level ($F = 4.37$, $df = 1$, $p < .05$). These results indicate that the effect of set is opposite in direction for a stigmatized as compared to a "normal" person. While for the former, a set to understand mitigates the severity of punishment administered, for the latter it increases the severity. An additional analysis was done in which the JN-group punishment scores were compared to the punishment scores of *Ss* in the other three groups. The *z* values yielded by the Mann-Whitney test were, respectively, 2.41 ($p < .01$) and 1.75 ($p < .10$) for the first and second series. Thus, a person is punished less when he is perceived as having had a typical childhood and being normal than when he is seen as having had a poor childhood or as having been mentally ill.

TABLE 1

MEAN INTENSITY AND DURATION OF SHOCK ADMINISTERED BY EACH GROUP FOR THE FIRST AND SECOND SERIES OF BUTTON PRESSES

Group	First series	Second series
Intensity		
JN	2.4	4.2
JS	3.3	5.1
UN	2.8	4.6
US	3.0	4.5
Duration		
JN	56.5	66.0
JS	69.5	88.4
UN	80.5	116.6
US	65.8	75.5

As indicated, the two components of the punishment index, intensity and duration of shock, were individually analyzed. Analyses of shock intensity revealed no significant main effects or interactions for either series. The results of the analyses of shock duration were similar to those of the punishment scores. The main effects of set and stigma were not significant for either series, and the interaction effect for the four-press series was also not close to statistical significance. However, the interaction *F* for the second series was only slightly smaller than that for the punishment score ($F = 4.27$, $df = 1$) and is statistically significant ($p < .05$). Also, as for the punishment score, the JN-group *Ss* receive shocks of shorter duration than *Ss* in the other groups. The Mann-Whitney *U* test *z* values were, for the first and second series, respectively, 2.0 ($p < .05$) and 1.75 ($p < .10$). This suggests that the punishment-score differences are primarily due to differences in the duration of the shock and are rather slightly enhanced by the intensity scores.

The 10 questionnaire items were analyzed by means of nonparametric tests to avoid making questionable assumptions about the measures.[3] These analyses also made it possible to retain all 11 *S*s in the JN group. The outcome revealed rather clearly that believing a person to be stigmatized leads to a less favorable evaluation of that person and, moreover, this evaluation is not seemingly changed by set. One item asked *S*s to indicate how likable they had found the confederate to be by choosing one of the five alternatives from very likable to not likable at all. A chi-square test revealed that the stigmatized person was significantly less liked than the normal ($\chi^2 = 4.26$, $df = 1$, $p < .05$). There was, surprisingly, a trend for the confederate to be less well-liked in the US than the JS condition. Analysis of another item showed that *S*s preferred the normal to the stigmatized partner if required to work on future similar tasks ($\chi^2 = 7.58$, $df = 2$, $p < .05$). They also estimated their own team performance as poorer relative to other teams when their partner was perceived as stigmatized ($\chi^2 = 14.3$, $df = 2$, $p < .001$).

The clearest influence of set was to change the feelings expressed toward the "normal" partner. The confederate was consistently evaluated more favorably in the JN condition than in the UN condition. As is indicated by the results already cited, the JN group also evaluated their partner more favorably than either the JS or US groups. One of the items indicating this requested *S*s to report how good a learner they thought their partner was. The JN group described their partner as significantly better than *S*s in the other groups combined ($\chi^2 = 4.11$, $df = 1$, $p < .05$). There were no apparent differences among the other groups. Another item for which significant set differences were found was the one calling for an estimation of adequacy of team performance. The JN group estimated their team performance as significantly better than *S*s in the other three groups combined ($\chi^2 = 5.21$, $df = 1$, $p < .025$). In addition, each of the remaining items which permitted expression of favorability showed a trend for the confederate to be most favorably evaluated in the JN condition.

There were no significant group differences for any of the remaining six items. For the two which requested an estimation of the confederate's intelligence and the degree of concentration he gave to the task, prior research suggests no differences were to be expected (Farina & Ring, 1965). Two other items asked for an evaluation of the shock administered. One required an estimation of the average intensity setting used and the second sought a judgment of how painful the shock was. The similarity of responses obtained across groups seems neither inconsistent with other results nor terribly surprising. It is also not surprising that there were no group differences for an item asking *S*s to estimate how effective they were as teachers since, in fact, they were all made to be equally effective in their "teaching." For the sixth and final item, which asked that primary blame for any failures be ascribed to either partner or self, differences had been expected. Both consideration and prior results (Farina & Ring, 1965) suggested the stigmatized partner would more frequently be blamed for joint failures than a normal co-worker. While there were directional trends consistent with this expectation, these were small and not significant.

DISCUSSION

It has been shown that stigma and the extent to which a person is held responsible for his stigma play a significant role in interpersonal behaviors. They clearly influence the amount of pain inflicted upon someone. The major question of determining if set is a variable which mitigates the unfavorable consequences of stigma is only partially answered, however. It was found, as had been expected, that a bad childhood experience tempers the harshness of the treatment accorded a person believed to be mentally ill. But, it was also found that this same experience has an opposite and extremely strong effect for the normal. The most painful shocks administered are given to the confederate when he presents himself as normal and as having had a bad childhood. This suggests that a bad childhood experience is itself stigmatizing and the questionnaire results amply confirm this. The confederate was as unfavorably evaluated when perceived as normal but having lived through a bad childhood period as when he presented himself as mentally ill. It may be that when a person has had a deviant upbringing he is perceived as mentally ill even if he claims good adjustment or perhaps any deviant experience is stigmatizing. Whatever the reason, the stigmatizing effects of an atypical childhood bring to light an interesting and unexpected phenomenon, although this also complicates the problem of determining the role of set in stigmatizing conditions.

---

[3] Whenever possible, chi-square tests were employed making use of each cell resulting from the choices of item alternatives. However, in accordance with the suggestion by Siegel (1956, p. 110), when the expected frequencies were too small median tests were used.

What the results clearly show is that a normal and rather typical person is treated more favorably than someone who deviates from this norm. Whether this departure takes the form of poor adult adjustment or bad childhood experiences, the deviant is dealt with in a harsher manner. The nature of the process accounting for this differential treatment is not entirely clear. Something about its characteristics is suggested by comparing the results obtained for the two components of the punishment index, however. Whereas intensity of shock did not differentiate the groups, the duration measure yielded a pattern of differences quite similar to that of the punishment index itself. Thus, as was pointed out in the Results section, the differences in amount of pain inflicted seem primarily due to the use of longer rather than more intense shocks. This suggests that when someone maltreats a stigmatized individual he may not be fully aware of what he is doing since no information is given as to how long the shock was administered but the intensity is selected and hence clearly known. Perhaps this abuse reflects a lessened concern about hurting someone when he is not completely normal. Or, possibly, the results are due to the perception of the stigmatized person as more incompetent and dull with this leading to the use of more painful shock in order to penetrate this dullness. In any case, these results are consistent with Goffman's (1963) observation that "we believe the person with a stigma is not quite human [p. 5]."

The results obtained with the questionnaire are highly consistent with prior studies. A stigmatizing condition causes the afflicted individual to be perceived as doing a poorer job and the joint performance is judged to be less adequate. This is after interaction with the stigmatized person and in the absence of any objective bases for these feelings, since the performance was identical for all conditions. Congruent with this are the findings that people seek to avoid future contact with a blemished individual and express dislike for him. In the case of the mentally ill person, set did not change these unfavorable evaluations, although it mitigated the harshness of the treatment accorded him. This finding carries the interesting implication that people can be made to behave more favorably toward stigmatized others while retaining their feelings of contempt and aversion. The reluctance to be the recipient of the traditional Christmas basket suggests this phenomenon may have been familiar to some classes of people a long time ago.

In the process of doing this study two kinds of questions were raised which deserve attention. One of these is why the report of a bad childhood experience should have such unfavorable interpersonal consequences. It seems more plausible to expect that the normal person would be given greater credit for having overcome this misfortune. Instead of this, he is disliked and mistreated as much as the mentally ill person. It may be that any atypicality is stigmatizing but a process more complex than this is suspected, such as perceiving him not only as maladjusted but as lying about his condition as well. The second kind of question is about the nature of the process which leads to the unfavorable treatment and evaluation of the stigmatized. In particular, the suggestion implied by the results that full awareness of how a person is behaving toward the mentally ill reduces the harshness of the behavior seems worth investigating. Experiments which would shed light on these two questions are not difficult to design.

REFERENCES

COHEN, J., & STRUENING, E. L. Opinions about mental illness in the personnel of two large mental hospitals. *Journal of Abnormal and Social Psychology*, 1962, **64**, 349–360.

COHEN, J., & STRUENING, E. L. Opinions about mental health: Hospital social atmosphere profiles and their relevance to effectiveness. *Journal of Consulting Psychology,* 1964, **28**, 291–298.

ELLSWORTH, R. B. A behavioral study of staff attitudes toward mental illness. *Journal of Abnormal Psychology,* 1965, **70**, 194–200.

FARINA, A., & RING, K. The influence of perceived mental illness on interpersonal relations. *Journal of Abnormal Psychology,* 1965, **70**, 47–51.

GOFFMAN, E. *Stigma.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

HASTORF, A. H., RICHARDSON, S. A., & DORNBUSCH, S. M. The problem of relevance in the study of person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford University Press, 1958. Pp. 54–62.

JONES, E. E., HESTER, S. L., FARINA, A., & DAVIS, K. E. Reactions to unfavorable personal evaluations as a function of the evaluator's perceived adjustment. *Journal of Abnormal and Social Psychology,* 1959, **59**, 363–370.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford University Press, 1958. Pp. 151–178.

JONES, N. F., & KAHN, M. W. Patient attitudes as related to social class and other variables concerned with hospitalization. *Journal of Consulting Psychology,* 1964, **28**, 403–408.

MANIS, M., HOUTS, P. S., & BLAKE, J. B. Beliefs about mental illness as a function of psychiatric status and psychiatric hospitalization. *Journal of Abnormal and Social Psychology,* 1963, **67**, 226–233.

NUNNALLY, J. C., JR. *Popular conceptions of mental health.* New York: Holt, Rinehart & Winston, 1961.

SHRAUGER, S., & ALTROCCHI, J. The personality of the perceiver as a factor in person perception. *Psychological Bulletin,* 1964, **62**, 289–308.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

STRUENING, E. L., & COHEN, J. Factorial invariance and other psychometric characteristics of five opinions about mental illness factors. *Educational and Psychological Measurement,* 1963, **23**, 289–298.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received November 1, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 429–431

FAILURE OF THE CONCEPTUAL STYLES TEST TO DISCRIMINATE NORMAL AND HIGHLY IMPULSIVE CHILDREN[1]

RHODA LEE FISHER

Syracuse Scholastic Rehabilitation Program, Syracuse, New York

2 cognitive dispositions, inferential and analytic, and their relationship to impulsive school behavior were investigated. The relationship of the dispositions to the Fisher-Cleveland Barrier score was also tested. The Kagan Conceptual Styles Test (CST) and the Rorschach ink blots were administered to 46 extremely impulsive boys in special 4th-grade school classrooms. The CST was not able to distinguish these impulsive children from Kagan's equivalent normal sample or from another normal group. As anticipated, no significant relationships existed between the Barrier score and either of the scores derived from the CST.

In a recent monograph, Kagan, Rosman, Day, Albert, and Phillips (1964) conclude that there are "two basic cognitive dispositions" which determine how external experiences are analyzed, classified, and even which "rules" are used in sifting the information being received by the observer. One disposition, the analytic, involves superior problem-solving performance, superior ability to attend, superior ability to delay, and superior ability in finding identical elements in what seems to be two quite different stimuli. To illustrate the last of these characteristics, one may take the case of a subject who is simultaneously presented with pictures of a watch, a man, and a ruler. The analytic subject would group the ruler and the watch together because "They both have numbers." The second basic disposition is the inferential approach. This categorization is based on functional relationships. Given the same three pictures: a watch, a man, and a ruler, the subject taking an inferential approach would group the watch with the man, saying, "The man wears a watch."

A technique developed by Kagan et al. to measure analytic versus inferential dispositions was called the Conceptual Styles Test (CST). Incidentally, the higher one disposition score is, the lower the other. Various behaviors were reported to be related to the two conceptual styles. Children with high analytic scores were termed "reflective" and those with low scores "impulsive." The impulsive group was described as "likely to display momentary lapses of attention during involvement in a school task: they were more likely to look out the window, gaze at a peer or orient to a sound during a period when they were working at an academic task." The reflective children were described as "capable of greater concentration; they showed less distractibility during conceptual tasks [Kagan et al., 1964, pp. 29–30]." Observations of the kinds of behaviors displayed by analytic versus inferential children during free play on the playground were consistent with the classroom behavior observations. Overall, the analytic children emerged as less impulsive and more carefully thoughtful in the way they arrived at decisions.

It was the main purpose of the present study to compare the analytic and inferential scores of a group of children known for their extreme impulsive behaviors with the scores of normal children. In this way Kagan's view could be evaluated, that is, that impulsivity and the inferential style of thinking are positively related. If the basic logic of the CST score is correct, the inferential scores of a group of children who are so impulsive that they are unable to adjust to the school classroom should be higher than those of a normal control group.

A second purpose of the study was to examine the relationship of the CST analytic score to a body-image boundary dimension developed by Fisher and Cleveland (1958)

---

[1] This study was partially supported by National Institute of Mental Health Grant MH-01475.

which Kagan et al. (1964) explicitly state should be related to it. The Fisher-Cleveland score is a score based on Rorschach percepts. Each percept given in response to the ink blots is scored in terms of its containing or bounding aspects. Fisher and Cleveland assume that the amount of boundedness in Rorschach percepts is a representation of the individual's own boundary articulateness. Kagan found that a score based on the number of well-articulated percepts produced when responding to ink blots was positively correlated with CST Analytic. With this perspective, he suggested that the Fisher-Cleveland index, which is based on degree of definiteness ascribed to the peripheries of ink-blot percepts, should also be positively correlated with CST Analytic. Indeed, he implied that the Fisher-Cleveland score could be better interpreted as a measure of cognitive vagueness than as an index of the clarity of one's body-image boundaries.

A third aim was to demonstrate that the Barrier score of impulsive children would be lower than the Barrier score for normal children. It has been demonstrated that the Barrier score is positively correlated with ability to deal with stress (Fisher & Cleveland, 1958), negatively correlated with delinquency (Megargee, 1965), and negatively correlated with gross schizophrenic disorganization (Fisher, 1963). One would assume, then, that a group of normal children who adapt adequately in school and who are able to control their impulses would differ in their Barrier characteristics from a group having difficulty in impulse control and adjusting to school demands.

PROCEDURE

The following tests were administered to each subject: (*a*) the CST, and (*b*) the Rorschach ink-blot series.

The standard series of Rorschach plates was administered individually, and each child was asked to give two responses to each of the 10 plates. The responses were categorized according to the Fisher-Cleveland Barrier scoring procedures. The Barrier score

> represents an evaluation of the degree to which the individual pictures the boundaries of his body as definite and firm as contrasted to indefinite and weak. The scoring of an individual's Rorschach protocol for this variable is based on a count of the number of percepts in which unusual covering, containing, decorative, or protective qualities are ascribed to the bounding peripheries. Such responses as the following would exemplify the category: "cave with rocky walls," "knight in armor," "animal with striped fur," "fort with cement walls" . . . The larger the Barrier score the greater is the definiteness of the body image boundaries [Fisher & Fisher, 1959, p. 391].

The scoring system has been shown to have high objectivity. Furthermore, adequate test-retest reliability has been demonstrated (Fisher, 1963).

Subjects

Ninety-two subjects from the Syracuse Public Schools were studied. Forty-six were white boys referred from the elementary schools because of their extreme impulsivity and lack of control. The group was obtained from a special referral service maintained by the Syracuse Public School system for children whom the regular classroom teacher considered too impulsive, too distractible to remain in her regular classroom. Their distractibility level was so high that it interfered with the instruction of the class as a whole. They are described by their teachers as extremely restless, impulsive, and capricious in response. Examples of the kinds of impulsive behaviors displayed are as follows: jumping in and out of seats, suddenly darting around the room, hitting other children without apparent reason, running away from the playground during recess periods. The mean age for the disturbed impulsive group was 9.4 years ($\sigma = 1.3$). All children included in this group were of at least average IQ.[2] The primary control subjects with whom the disturbed impulsive subjects were to be compared were obtained from the norm groups described in the Kagan et al. monograph.

A normal group was also collected which came from classrooms of the same public school as the disturbed subjects. The mean age of this group was 11.2 years ($\sigma = .63$). The children comprising this normal sample were of at least normal intelligence. The prime reason for collecting this group was to examine the relationship between Barrier and CST scores. It was further possible, by extracting a subgroup which was matched in age with a subgroup from the disturbed impulsive sample, to determine if the CST scores would differ when the disturbed subjects were compared with still another norm group.

RESULTS

The Conceptual Styles Test

The original norms presented by Kagan for the CST state that the mean analytic score for fourth-grade boys is 7.5. A comparison between the Kagan norms and the impulsive disturbed boys in the present sample who were of fourth-grade level, indicated no difference between them. The mean score for the present sample was 7.9 ($\sigma = 4.4$) which, if anything, is slightly higher than Kagan's norm of 7.5.

---

[2] To be admitted to a special class the disturbed child must not be of less than average intelligence.

A second analysis was done in which all boys 10–11.5 in the normal group ($N = 24$) collected in the present study were compared with all boys of equivalent age in the disturbed impulsive sample ($N = 23$). The mean analytic score for the normal control group was 8.7 and that for the disturbed impulsive group was 8.4. The difference between these two is negligible.

Fisher-Cleveland Barrier Score

The Barrier score, when correlated with the analytic scores, showed no significant relationships. Thus, for the disturbed impulsive group of children a correlation of $-.08$ ($N = 46$) was obtained between Barrier score and analytic score. In the total normal group the Barrier score correlated only $-.01$ ($N = 45$) with the analytic score.

The mean Barrier score for the disturbed impulsive group ($N = 46$) in the present study was 4.1, whereas the mean for the total normal group ($N = 45$) was 5.3 ($t = 2.8$, $p < .01$). A second analysis of Barrier score, holding age constant, was also obtained by using only subjects of 10–11.5 years in each group. The mean Barrier score for the disturbed impulsive group ($N = 23$) was 3.9; and for the normal group ($N = 24$), 5.4 ($t = 2.5$, $p < .05$).

DISCUSSION

From the evidence presented in the present paper it is clear that the CST analytic score does not seem to be able to differentiate a highly impulsive acting-out group of children from a normal sample. Obviously, CST inferential and impulsivity cannot be regarded as synonymous. Level of distractibility in the present school-problem group is far beyond the acceptable range of school behavior. If the CST analytic score cannot differentiate a group presenting such a behavior extreme from a group of normal school children, what is it that the "analytic disposition" actually taps? Perhaps a reinterpretation of the meaning of the original findings described by Kagan is in order. Kagan's typical CST analytic child was one who took more time, put off rewards, was concerned about his language, and carefully appraised visual stimuli. Possibly such a child is not so much "reflective" as anxiously indecisive. Perhaps one is dealing with hesitancy rather than ability to delay. Children who take time to make up their minds may do so because of anxiety.

The present study has demonstrated also that the cognitive dispositions defined by Kagan are, contrary to his prediction, in no way related to the Barrier score developed by Fisher and Cleveland. Apparently, the Barrier score cannot be defined in the simple fashion that Kagan suggested in his monograph. Also, the difference between the normal and disturbed groups in Barrier score is supportive of previous work with the Barrier score in which similar results have been found when comparing normal and other types of disturbed samples (Fisher & Cleveland, 1958). It is also supportive of a recent study by Megargee (1965) in which delinquents proved to have lower Barrier scores than nondelinquents. Boundary definiteness seems to be meaningfully and reliably related to erratic, impulsive acting-out.

REFERENCES

FISHER, S. A further appraisal of the body boundary concept. *Journal of Consulting Psychology,* 1963, 27, 62–74.

FISHER, S., & CLEVELAND, S. *Body image and personality.* New York: Van Nostrand, 1958.

FISHER, S., & FISHER, R. A developmental analysis of some body image and body reactivity dimensions. *Child Development,* 1959, **30**, 389–402.

KAGAN, J., ROSMAN, B., DAY, D., ALBERT, J., & PHILLIPS, W. Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs,* 1964, **78**(1, Whole No. 578).

MEGARGEE, E. I. Relation between barrier scores and aggressive behavior. *Journal of Abnormal Psychology,* 1965, **70**, 307–311.

(Received November 1, 1965)

ent purposes as well as his own, his proposed solution is in more direct contradiction to the present one than is the received system. For cathexes, bound, free, hyper-, or counter-, are drained of psychological meaning, hence belong to a different universe of discourse than what is here called psychoanalytic psychology. Effectance, a sense of effectiveness, contains intelligible meaning, hence belongs within our chosen universe of discourse. The phenomena covered by effectance will be shown encompassed by traditional principles of psychoanalytic psychology. In particular, there is a clear psychological meaning to the assertion that an impulse does in the normal course of events often lead to or precipitate its own control, without invocation of any (nonpsychological) concept of energy.

The purpose of the present essay is to demonstrate three principles, or perhaps areas, of psychoanalytic psychology. These principles are distinctively if not quite exclusively human and distinctively if not quite exclusively psychoanalytic. Major aspects of ego development can be accounted for by the principles as well as the traditional subject matter of psychoanalysis (here following Fenichel). They are all found, explicitly or implicitly, in the writings of Fenichel, Rapaport, and many others, but originally in the writings of Freud. They are not reducible to or derivable from biological or tension-reduction or pleasure principles (here following Hendrick and White). While precise formulation of the principles is not easy, one can say where they are to be looked for. The three areas are those of the dynamic unconscious, the plasticity of the interpersonal drives, and mastery of experience through reversal of voice. The question to be raised is: Are any other major dynamic principles necessary? Are additional concepts and principles ad hoc and supernumerary?

Psychoanalytic ideas by themselves cannot suffice to account for psychological observations but must be supplemented by ordinary terms and observations from the common domain. That a drive seeks gratification is such a notion from the common domain, as is also the observation that any gratification, once tasted, is reluctantly surrendered. Structuralization, that what one experiences determines what one becomes, is also a general psychological principle. Cathexis and its variants, in contrast, are not common sense notions and cannot justly be invoked to patch holes in the theory. The three following propositions are what Hartmann and Kris (1945) have called dynamic propositions. No corresponding simplification of what they call genetic propositions is here attempted. Rather, it will be shown that some well-known genetic facts can be accounted for by the three dynamic principles.

THE DYNAMIC UNCONSCIOUS

The principle of the dynamic unconscious states, certainly, that unconscious ideas influence behavior in somewhat the same way that conscious ideas do (Freud, 1900b, p. 608),[2] but that is too weak a formulation. That the most important determinants of behavior are unconscious is too strong, though it was once popularly believed to be the psychoanalytic view. One version might be that a given idea influences behavior more strongly when it is unconscious than after it is made conscious. The most elegant version, though it must be qualified to make it strictly true, states: Both conscious and unconscious ideas affect behavior, but the consequences of behavior affect only its conscious sources. In this formulation preconscious ideas are obviously grouped with conscious ones.

Not encompassed in the principle are the appearance in consciousness of derivatives of the unconscious, the institution of censorship, and other complications. If this version held exactly, therapy would be impossible. The principle is a schematic version of the short-run relation of conscious and unconscious determinants to behavior. Behavior originating in the unconscious is not stamped out by ordinary rewards or consequences; hence, this is the Law of No Effect.

PLASTICITY OF THE INTERPERSONAL DRIVES

The second principle can be stated that those drives that normally take other people as their object, that is, the sexual and aggressive drives, can be satisfied by a wide variety of acts as well as objects. Although psychoanalysts now generally agree that the drives for their study are the sexual and aggressive ones, they are not always clear why it is these needs rather than such imperative and instinctual drives as thirst, oxygen need, and sleep. The latter drives do not necessarily or normally express themselves in relation to other people, they are resistant to delay, and they do not permit substitution to any extent. The importance for psychoanalytic psychology of sex and aggression results from their plasticity and their interpersonal character together. The interpersonal nature of the impulses demands their socialization; their plasticity makes it possible. The plasticity and the socializing pressures together lead to psychic elaboration. Socializing pressures at times take the form of taboo. There is thus an integral, logical reason for psychoanalytic psychology to concern itself with the sexual and aggressive drives, precisely the ones around which there are taboos. They are not the only drives, nor the most life-important ones, but they are the social ones and the psychically elaborated ones. The principle must be broadened slightly to cover the fact that in infancy, for obvious reasons, the baby takes his own person as object, or more properly, the baby and the object-person(s) are undifferentiated at first. The sex drives are manifold and physiological in origin, hence arise out of the range of psychology proper. Although the biological origins of the sex drive are consistent with this formulation, though outside its scope, there is an alternative way of thinking of the pregenital phases of psychosexual development, to wit, as the sexualization and aggressivization of relatively nonplastic drives, such as hunger and defecation. The most imperative drive, the need for oxygen, seems to be least often sexualized. Rapaport (1960b, p. 47, p. 129) speculated that emphasis on the sex drives was not necessarily a permanent feature of psychoanalysis as a system, a view Freud did not share. The present formulation supports the orthodox emphasis.

---

[2] References to the works of Freud are to the Standard Edition. For clarity of historical reference, dates given in the text refer to date of first publication in these and a few other cases. For such references the date of original publication appears directly after the author's name.

Freud's early psychological writings are cast in the format of the plasticity principle: dreams, slips, jokes, and symptoms are disguised expressions of unconscious or suppressed wishes (though jokes are closely related to play, a subject he treated at length some years later). The term *wish* is entirely within the domain of psychology, that is, of meaning, but in the reductionistic fervor of later years it passed from the technical vocabulary, being replaced by the ambiguously biological terms *instinct*, *drive*, and *instinctual drive*. Current usage permits substitution of *impulse*, which will be used where possible as a compromise between (the biological or pseudobiological) drive and (the truly psychological) wish.

The ontogeny of sexual drives constitutes one of the major observational findings of psychoanalysis. Although it has been stated (Waelder, 1960) that the aggressive drive undergoes no similar developmental transformation, perhaps a slight shift of focus will reveal ego development as the analogue within the sphere of aggression, broadly conceived, of psychosexual development within the sphere of libido.

MASTERY THROUGH REVERSAL OF VOICE

The third principle, which may be called the principle of the converse consequence, states that experience is mastered by repeating in an active role what one has previously experienced in a passive role: one must do what one has suffered. What alter has done to ego, ego must do to alter. But thanks to the plasticity of drives, I need not necessarily do to you exactly what you have done to me, nor need the object of my actions be the same person who has done something to me. In infancy and early childhood one may assume a large measure of substitutability of people and animals, animate and inanimate objects, nor is this propensity for substituting new objects ever lost.

The principle of the converse consequence obviously has a driven, a drive-like quality; hence, the principle of drive plasticity is used to account for its vicissitudes. While the sex drive is of biological origin, the aggressive impulse is more obscure in origin. An enormous simplification is achieved by putting the principle of the converse consequence as a solution to the conundrum of the origin of the aggressive drive. Everyone begins life in a helpless and passive position. Each step of growth in the direction of active mastery precipitates one into new experiences that must in turn be mastered. This principle alone thus can account for a bottomless reservoir of need for mastery. The need for mastery can be thought of as identical with the aggressive drive if one interprets aggression broadly to include exploration, play, and other aspects of what White calls effectance. If the term aggression is limited to hostility, then aggressive impulses are only part of what is included in this principle, but the principle still includes all of what has been called the aggressive drive.

Freud's clearest exposition of the principle of the converse consequence is in *Beyond the Pleasure Principle,* first published in 1920. Children's play is given as an example of phenomena requiring for their account the pleasure principle but also an "instinct for mastery [Freud, 1920, p. 16]." In throwing away his toys, playing "gone" with them, the child mastered by repeating in an active part what he had experienced passively when his mother left him periodically; the game also expressed his impulse to revenge himself on his mother for leaving him. Similarly, when children suffer indignities at the hands of physicians and dentists, they soon subject a playmate to a version of the experience, again mastering by changing from passive to active role, revenging themselves on a substitute, and also expressing "a wish that dominates them the whole time—the wish to be grown-up and to be able to do what grown-up people do [Freud, 1920, p. 17]." Apparently Freud meant that mastery through reversal of voice and the pleasure principle, expressed in terms of the wish to be grown-up and the wish for revenge, are partially competing explanations, but that both are required to account for play.

To speak of a child's wishes to be grown-up and to take revenge is to reveal the vacuity of the pleasure principle. The operative principles are those that account for the origin of such wishes, not for the fact that they seek gratification, which says no more than that they exist. Are both these wishes not expressions of desire for mastery through changing from a passive to an active role? If so, the two interpretations, that in terms of pleasure principle and that in terms of reversal of voice, reduce to aspects of a single one; the impulse to master operates like a drive and seeks gratification in many alternative ways.

Beyond the Pleasure Principle is commonly stated to be the source of what is here called the principle of the converse consequence, but that statement is not accurate. The principle was part of psychoanalytic thinking from the beginning. In *The Interpretation of Dreams* there are a number of references to reversal of voice as a principle of disguise. More germane to the present topic, Freud explained one of his dreams saying, "Turning the tables on His Excellency with a vengeance! He had refused to appoint me *professor extraordinarius* and I had retaliated in the dream by stepping into his shoes [Freud, 1900a, p. 193]." "Turning the tables" is a nice paraphrase of the converse consequence; "turning the tables with a vengeance" expresses the frequent but not invariable association of this principle with hostile aggressive impulses.

In *Civilization and its Discontents* Freud (1930) reaffirmed the primitive and instinctual nature of the aggressive drive and added that the institution of conscience derives entirely from transformation of the aggressive drive rather than from libido (hence the above statement that ego development is to the aggressive drive what psychosexual development is to libido). The development of conscience is a central problem in the institution of culture. Omitting Freud's speculations about murder of the primal father, his account of the origins of conscience and of a sense of guilt is as follows.

The child is restrained in expression of his impulses by external authority (Freud's insistence that it be the father is long ago and far away). The child wishes to take vengeance on his parent for imposing restraint, but this is not permitted; besides, it would endanger the parent's love, which he needs. He therefore masters this situation by taking the vengeance on a part of himself, namely, the errant impulses, that he would like to have taken on the parent. Moreover, by means of identification he assumes the role of the parent, treating part of himself, that is, his impulses, as his parent treated him or them. By both of these means he establishes control of his impulses. The former particularly accounts for the observation that the child may be much harsher with himself than his parents would have him. The child thus reproduces within himself the relation of obedience to authority, as a means of mastering an otherwise frustrating situation. The internalized authority is the superego, which watches over the ego as the parent watched over the child.

Returning now to the principle of the converse consequence, what he has done to me I must do to him, remembering that Freud wrote German, and making use of the possibility of substituting a victim for our vengeance, we have: What he (*er*) has done to me (*mich*), I (*ich*) must do to it (*es*). (A grammatical lapse must be permitted in German to make a point.) Translating back into English, substituting "parent" for "he," we have: What my parent has done to me, my ego must do to my id.

This formulation differs in this much from that of *Civilization and its Discontents*. Freud speaks of the superego as exercising authority over the ego, whereas our version depicts ego as controlling id, that is, impulse. In phenomenological terms these descriptions are not entirely different. In either case the child has become internally differentiated using an interpersonal relation both as model and as impetus. To the extent that the differentiation of ego from impulse is something other than the differentiation of superego from ego, they are successive stages in the formation of conscience, indeed, in the development of the ego. In any case, Freud's account is compatible with explanation by successive applications of the principle of the converse consequence.

There must be some stage at which the child experiences his impulses as controlling him. Again, this experience can be mastered by his ego controlling his impulses as it has been controlled by them. This is another and coordinate version of how the principle of mastery through reversal of voice accounts for an impulse being enlisted in its own control.

But at this point we have solved, or proposed a solution to, one of the most baffling theoretical problems of psychoanalysis. Psychoanalytic theorists from Freud and Breuer on have approached the problem of how an impulse contributes to its own control by postulating as new entities bound cathexis and countercathexis, sometimes also hypercathexis. But why is cathexis bound? How does cathexis become transformed to countercathexis? The principles were already at hand without multiplying new constructs. One is controlled, simultaneously or in turn, by one's impulses and one's parents, and one therefore needs to control something in return, if not one's parents, then one's impulses.

In some formulations identification is invoked to explain this process. But is identification an explanation? Or, rather, does the principle of the converse consequence have identification as one of its manifestations? The latter would seem to be the simpler version. At this point an additional principle has been assumed, namely, that experiences create structures, that is, permanent or semipermanent changes in the person. What one does becomes what one is. This assumption must be made in psychoanalytic theory, but it is a general psychological principle, part of the common domain, rather than a specifically psychoanalytic principle. How else would learning of any sort be possible?

Another classical monograph that continues the line of thought of *Beyond the Pleasure Principle* is Anna Freud's (1936) essay on *The Ego and the Mechanisms of Defence.* One of the best known and most influential concepts of this monograph is her exposition of the mechanism of identification with the aggressor, which she conceives of as a way station in the development of the superego, a precursor of and substitute for guilt feelings. True morality requires not only taking over the adult's criticism but turning it into self-criticism. The underlying dynamic principle, she makes clear, is the change from passive to active as a means of assimilating unpleasant or traumatic episodes. Another defense mechanism, that of "altruistic surrender," common in parents and in some other persons, she also explains in terms of the principle of mastery through reversal of voice.

Let us recapitulate briefly. One masters reality by turning the tables, by actively repeating what one has passively experienced. Since this principle governs the relation to reality, it is a core of the ego's functioning. But it is also a governing principle in the most essential stages of the formation and development of the ego, namely, the initial separation of ego from impulse, the identification with the aggressor, or, more broadly, with authority, and the internalization of the dictates of authority in the form of conscience and self-criticism. If one conceives of altruism as a still further stage of ego development, it becomes further evidence that ego development proceeds by application of the principle of the converse consequence (hence, using hyperbole, by transformation of the "aggressive drive").

Clearly, the proposed three principles account for much of normal ego functioning and of ego development. Adequately to represent psychoneurosis, which is due to inner conflict, one must allow for a short circuit in the elaboration of interpersonal impulses. Socialization, the pressure for deformation of the impulse, must arise within the individual himself. How that happens, how the interpersonal clash becomes the model and the occasion for an intrapersonal clash in the normal course of ego development, has just been described as a consequence of the principles already assumed. The extraanalytic assumption of structuralization permits us to accommodate the fact that ego mediates both the socializing impress of the environment—and at just this point inner conflict, hence neurosis, becomes possible—and also the motility by means of which impulses attain their objects.

ANXIETY AND SEXUAL CONFLICT

A mature version of one aspect of Freud's views on pathology can be found in *Inhibition, Symptoms and Anxiety* (1926). "The situation [an infant] regards as a 'danger' and against which it wants to be safeguarded is that of non-satisfaction, of a *growing tension due to need,* against which it is helpless [Freud, 1926, p. 137]." The infant comes to fear loss of mother—since he does not know of the permanence of objects, her temporary disappearance is experienced as loss—in advance of being overcome by excess stimulation. "This change constitutes a first great step forward in the provision made by the infant for its self-preservation, and at the same time represents a transition from the automatic and involuntary fresh appearance of anxiety to the intentional reproduction of anxiety as a signal of danger [Freud, 1926, p. 138]."

> Let us call a situation of helplessness of this kind that has been actually experienced a *traumatic situation.* . . . A danger-situation is a recognized, remembered, expected situation of helplessness. Anxiety is the original reaction to helplessness in the trauma and is reproduced later on in the danger-situation as a signal for help. The ego, which experienced the trauma passively, now repeats it actively in a weakened version, in the hope of being able itself to direct its course [Freud, 1926, pp. 166–167].

The later situations are most characteristically loss of mother, loss of parent's love, castration, and punishment from the superego. Freud traces briefly how each of these types of anxiety reproduces and represents a transformation of the original anxiety, in each case containing some element of helplessness in the face of mounting tension.

A modern analogy may add to and clarify an aspect of this exposition. One may think of the baby's experience as a message sent and his consequent actions as the corresponding message received. The transition from the passive to the active voice is in this analogy the channel of communication. The helplessness of the traumatic situation thus corresponds to an overloading of the channel of communication, so that no message can be sent. To put it simply, the baby cannot express all that he feels. What he cannot express constitutes the trauma and is repressed. Apparently there is here some prototype of repression, even prior to the differentiation of an ego. This residue of unassimilated experience thus becomes part of the unconscious, and, in accord with the first principle, forever sends messages to behavior without receiving a message in return.

Freud's discussion of the castration complex as growing directly out of the infant's fear of helplessness in the face of mounting tension offers a bridge from the principle of mastery through reversal of voice to problems of the evolution of sexual impulses, that is, psychosexual development. The attempt to explain the difficulties of female sexuality in terms of the need to shift the object of affection from mother to father is not convincing on Freud's own logic. For the first caretaker must be a withholder as well as a giver and hence is inevitably an object of ambivalence. A priori, shifting love to father might as well be a simplification of the girl's libidinal life as a complication (Freud, 1931). Much clearer is that the need to shift from one organ, the clitoris, to another, the vagina, offers an obstacle to the girl's harmonious development. But if the little girl fails to discover (as Freud believed) or only dimly perceives her vagina, how much more dimly must she perceive the possibility of gratifying discharge of sexual tension in the passive-receptive role. At the same time, the traumatic possibilities of helplessness in the face of tension are all too well known. Hence the impulse to master through activity, which governs the girl's ego functioning and development as much as the boy's, is in fixed opposition to her ultimately harmonious and gratifying sexual position. Here is a clear paradox in female sexuality, necessarily in force in childhood and never completely obsolete.

A boy's sexual development is simpler through being spared this paradox, but he experiences his own typical dangers. The ego is differentiated out of a primal unity, not only an ego-id matrix but also an ego-reality matrix (Loewald, 1951). Boundaries of self and of objects must necessarily be established simultaneously. This first stage of primitive unity and immediate gratification has been described both in terms of the infant feeling helpless and feeling omnipotent; yet these descriptions are not contradictory. Although one can hardly expect to describe the state with words, the infant may be said to be omnipotent in participating in the omnipotence of his environment, but impotent to the extent he is differentiated from it. The attraction of the passive-receptive gratification of the primal unity is hazardous to a girl's ego as well as to a boy's, since it implies not only helplessness but a threat of ego dissolution. On the whole, however, a girl can integrate remnants of this period into her later life, social as well as sexual, more easily than can a boy. The inevitable occurrence of the passive role or position in nonsexual sectors of life may become inordinately threatening to him. The bearers of the most modest authority, male or female, may become potential symbolic mutilators. Thus the high school boy may be disturbed by the requirement to do his homework, attend school regularly, or abide by ordinary family rules. The very concordance of his ego needs and his sexual drives may make a threat to one doubly intolerable.

THE ECONOMIC PROBLEM

White (1963) has criticized the received psychoanalytic version of ego development. A great simplification is achieved, he asserts, by postulating, in addition to the sexual and aggressive drives, independent ego energies that account for play, exploration, competence, reality testing, and other phenomena; this phenomenological view is simpler, more convincing, less tortured than the patchwork of drive-derivative accounts of ego development, invoking cathexis, anticathexis, bound cathexis, hypercathexis, neutralization and deneutralization, fusion and defusion of drives. Had he been more meticulous to couch his proposal entirely in psychological terms, avoiding all reductionistic ones like energy (as he might have without changing his argument in any essential), the difference between his view and the present one would be small, though not negligible.

The principle of the converse consequence has been shown in its favorable and benign operation to be the chief or perhaps the entire source and nature of what White calls independent ego energies, that is, of the impulses toward effectiveness and competence, as distinct from drives characterized by tension reduction, but it is also the origin of the aggressive impulses. The present form is simpler than White's version. He does not cope with the paradox that aggression has the quality of a drive though lacking the organic base of sex and other drives. He also avoids coping with the similarity of positive (effectance) and negative (hostile) aspects of aggression. On the other hand, the present version leaves one with the necessity to account for their difference. That both are aspects of what Hendrick termed the "instinct to master" does not lessen the need to account for the difference. At this point it appears one must have recourse to quantitative considerations. Where the child's experience is such that he can express his reaction to it adequately, where he can cope, abreact, or discharge, as one will, then he can play, explore, and acquire competence. Where the intensity of his experience or the limited possibility of reaction leaves the child as a channel of communication taxed or overwhelmed, he will range from anger to rage, from anxiety to repression and other pathological phenomena.

In psychoanalytic writings quantitative considerations are usually termed the economic view. The economic view serves two quite different functions, on the one hand being the essence of Freudian reductionism, on the other serving to replace a monistic principle of human functioning. Freud often expressed the view that with true scientific understanding, the qualitative or meaningful descriptions would be replaced by purely quantitative ones. But what happened was in at least one instance exactly the opposite:

> At one time I attached some importance to the view that what was used as a discharge of anxiety was the cathexis which had been withdrawn in the process of repression. To-day this seems to me of scarcely any interest. The reason for this is that whereas I formerly believed that anxiety invariably arose automatically by an economic process, my present conception of anxiety as a signal given by the ego in order to affect the pleasure-unpleasure agency does away with the necessity of considering the economic factor. Of course there is nothing to be said against the idea that it is precisely the energy that has been liberated by being withdrawn through repression which is used by the ego to arouse the affect; but it is no longer of any importance which portion of energy is employed for this purpose [Freud, 1926, p. 140].

In other words, when anxiety was understood as a signal, that is, as meaningful, the economic version was no longer of interest, exactly the reverse of the sequence anticipated by programmatic reductionism (which Freud nonetheless continued to advocate).

Even for those who disavow reductionism, who wish to remain psychologists concerned with meaningful units of behavior, the quantitative view cannot be ignored. Goldstein (1939) asserts that there must be just one principle of human conduct, for if there are two principles then there must be a superordinate principle to decide which one is in force at any moment, and it will be the superordinate principle that then becomes the one principle. For him this superordinate principle is self-actualization. Psychoanalysis does not uphold any such single principle to account for behavior. Rather, of the several forces brought to bear at any one time, it is their quantitative relations or, as Waelder (1960) phrases it, the parallelogram of forces, that must decide the outcome. At the present time virtually nothing has been accomplished by way of independent measurement of quantitative factors. Hence they invariably have been inferred from the outcome. We may conclude that while a quantitative assumption is needed to avoid the logic of a monistic psychology, its sole function is to ratify the observations. The quantitative assumption covers what we do not understand, and it loses interest as fast as we understand meanings.

SOME IMPLICATIONS

The principle of mastery through reversal of voice has been invoked by psychoanalysts beginning with Freud in a wide variety of contexts; yet Rapaport, who was thoroughly familiar with it (Rapaport, 1953), failed to include it in his systematization of psychoanalysis (Rapaport, 1960b) or in his exposition of psychoanalysis as a developmental psychology (Rapaport, 1960a). On the other hand, Hendrick (1942), who drew attention to the central importance of the "instinct to master," failed to state its specifically psychoanalytic version, that mastery is achieved by reversal of voice. The latter criticism applies also to White's (1963) "independent ego energies." In Erikson's (1950, ch. 6) writings, particularly in those concerning the autocosmic sphere, the microsphere, and the macrosphere, the principle is given its due as central to psychoanalytic thinking.

At least some people in some aspects of their lives arrive at a state where the principle is greatly modulated; they can stop fighting without striking the last blow. On the other hand, many adults do not outgrow the crude manifestations of the principle. There are unanswered questions concerning the scope and vicissitudes of the principle. One question is, what constitutes a blow? What kinds of experience require to be mastered? For an infant perhaps almost everything, for a well-functioning adult, much less. A general answer probably should be couched in terms of stage of ego development. A second question is, what will the person find acceptable as counterblows? Again the answer will depend on ego level as well as other attributes but seems not to have been worked out in general terms as yet.

At this point we are committed to what may look like a circularity. The ego develops as the person works to master his experience; yet what requires to be mastered and what is acceptable to him as effective in mastering depends on ego level. This need not be a circularity but rather a statement that the ego develops slowly and by its own logic.

The theory of ego development abstracted here from the writings of Freud and Anna Freud can be restated in terms and in a style of thought borrowed from Piaget. Interpersonal schemas serve as models for intrapersonal schemas, with frustration of (ordinarily) interpersonal impulses as impetus. Loewald (1960) specifically describes ego development in such terms, in a paper applying the same paradigm to psychoanalytic therapy, as both analogue and continuation of ego development. (The principle of the converse consequence is itself a schema, similar to that of the mechanism of undoing. Whether the schema originates from the formally similar schema of Newton's third law is beyond present scope. Certainly interpersonal schemas are not the first ones apprehended.)

The experiences that give rise to structures seem to be just those that become unconscious. According to the principle of the dynamic unconscious, however, it must be the other way round. What is unconscious is thereby rendered semipermanent, hence, by definition, becomes structural. Common sense requires a distinction between symptoms, on the one hand, and the formation of ego and establishment of conscience, on the other; indeed, common sense may require a sharper distinction than therapy can sustain. To sustain the distinction between symptom and structure, theory, like therapy, must have recourse to the social matrix for which Erikson has been an eloquent spokesman. That lies outside the scope of the present essay, together with many other questions, such as how consciously remembered experiences can participate in formation of ego structure, as they surely do.

There appear to be other theories of ego development, the monistic ego psychologies and the currently received psychoanalytic version. The monistic ego theories state that ego development occurs because man strives for mastery, competence, or self-actualization. These theories come close to saying no more than that ego development does indeed occur. Certainly they state too little to be accounted substantial theories.

The received psychoanalytic version (Hartmann, 1939; Rapaport, 1960b), by contrast, assumes too much. In the numerous variations of cathexis, neutralization, deneutralization, fusion, and defusion of drives, it postulates more entities and processes than require to be accounted for.

Thus the theory of ego development presented here, that interpersonal schemas become intrapersonal schemas, is not necessarily truer than the received psychoanalytic version or than the monistic ego psychologies, but it is much closer to qualifying technically as a scientific theory. It asserts something other than what can be observed more or less directly, but it does not postulate a set of entities more complex than the observations.

CONCLUSION

The three principles presented here have been characteristic of psychoanalytic thinking from its beginnings. The principle of the dynamic unconscious is presented clearly in *Studies on Hysteria* (Breuer & Freud, 1895). The principle of the plasticity of the interpersonal drives is anticipated there in the stress on sexual etiology and on the repression of sexual impulses where they are incompatible with the patient's moral sensibility or other aspects of the ego. The emphasis on abreaction and catharsis anticipates the principle of the converse consequence. The paradigm of the plasticity of the interpersonal drives has survived radical changes in doctrine about what the fundamental drives or impulses are. Freud's reductionistic theory has covered the same span of time as his psychology, beginning with his 1895 "Project for a Scientific Psychology" (Freud, 1954). The question is raised whether the three psychological principles suffice to cover also those phenomena usually described in terms of cathectic energy.

Obviously, this essay cannot answer by reconstructing the whole history and scope of psychoanalytic psychology. The three principles constitute something akin to a system of differential equations. The macroscopic phenomena of normal development, of pathology, of therapy and recovery are due to the simultaneous operation of all three principles rather than being classifiable under one or the other. Thus in ordinary life none of the principles need be visible in pure form.

The vast literature of psychoanalysis is not replaced but supplemented by an attempt to discover the simplest, most economical set of principles that can account for the phenomena. The ground rules for evolving such a set of principles have been that they should be distinctively human and distinctively psychoanalytic, hence not reducible to biological or tension-reduction principles, and hence also saving the phenomena to be explained. That is, they should be concerned with meanings, as psychoanalytic therapy is. (There has been also a covert assumption that what is meaningful can be stated in ordinary language, without neologisms.) No system of ideas can be complete without indefinitely much dependence on the common domain. That a drive seeks gratification is such an idea from the common domain. Psychoanalysis must be consistent with biological observations but it need not account for them.

The psychology abstracted by these rules from the writings of Freud is a true and magnificently economical scientific theory. Psychoanalysts have tended to take this theory for granted in their work and have extolled as theory Freud's notions about cathectic energy. But the energic theory is enormously complex, made increasingly so as its inadequacies are increasingly seen, and it neither simplifies nor saves the observed phenomena. We have asked: Are any psychoanalytic principles necessary other than those of the dynamic unconscious, the plasticity of the interpersonal drives, and the converse consequence? The answer will take years of thoughtful consideration of diverse topics. Indeed, one might ask for a more careful examination of what extraanalytic principles are needed to make a rigorous system, were it not that systematization quickly becomes as dreary as the disease it aims to cure. What has been presented here is an audacious *tour de force,* too sketchy to be a model of mind. But it is not too simple to be a model of a theory of mind. The question can now be restated: Had it not been intertwined with Freud's psychology, to which it is irrelevant, would Freud's cathexis theory even have survived as a scientific theory?

REFERENCES

BREUER, J., & FREUD, S. (Orig. publ. 1895) *Studies on hysteria.* Vol. 2. (Standard ed.) London: Hogarth Press, 1955.

ERIKSON, E. H. *Childhood and society.* New York: Norton, 1950.

FENICHEL, O. Early stages of ego development. *Imago,* 1937, 23, 243–269.

FINGARETTE, H. *The self in transformation.* New York: Basic Books, 1963.

FISHER, A. L. Freud and the image of man. *Insight,* 1963, 1, 13–26.

FLEW, A. Motives and the unconscious. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science.* Vol. 1. Minneapolis: University of Minnesota Press, 1956. Pp. 155–173.

FREUD, A. (Orig. publ. 1936) *The ego and the mechanism of defence.* New York: International Universities Press, 1946.

FREUD, S. (Orig. publ. 1900) *The interpretation of dreams.* Vol. 4. (Standard ed.) London: Hogarth Press, 1958. (a)

FREUD, S. (Orig. publ. 1900) *The interpretation of dreams.* Vol. 5. (Standard ed.) London: Hogarth Press, 1958. (b)

FREUD, S. (Orig. publ. 1920) *Beyond the pleasure principle.* Vol. 18. (Standard ed.) London: Hogarth Press, 1955. Pp. 7–64.

FREUD, S. (Orig. publ. 1926) *Inhibition, symptoms and anxiety.* Vol. 20. (Standard ed.) London: Hogarth Press, 1959. Pp. 75–174.

FREUD, S. (Orig. publ. 1930) *Civilization and its discontents.* Vol. 21. (Standard ed.) London: Hogarth Press, 1961. Pp. 64–145.

FREUD, S. (Orig. publ. 1931) *Female sexuality.* Vol. 21. (Standard ed.) London: Hogarth Press, 1961. Pp. 225–243.

FREUD, S. *The origins of psychoanalysis. Letters to Wilhelm Fliess, drafts and notes: 1887–1902.* New York: Basic Books, 1954.

GOLDSTEIN, K. *The organism.* New York: American Book, 1939.

HARTMANN, H. (Orig. publ. 1939) *Ego psychology and the problem of adaptation.* New York: International Universities Press, 1958.

HARTMANN, H., & KRIS, E. The genetic approach in psychoanalysis. *Psychoanalytic Study of the Child,* 1945, **1**, 11–30.

HENDRICK, I. Instinct and the ego during infancy. *Psychoanalytic Quarterly,* 1942, **11**, 33–58.

HILGARD, E. R. Impulsive versus realistic thinking: An examination of the distinction between primary and secondary processes in thought. *Psychological Bulletin,* 1962, **59**, 477–488.

HOLT, E. B. *The Freudian wish and its place in ethics.* New York: Holt, 1915.

HOLT, R. R. A critical examination of Freud's concept of bound and free cathexis. *Journal of the American Psychoanalytic Association,* 1962, **10**, 475–525.

LASHLEY, K. S., & COLBY, K. M. An exchange of views on psychic energy and psychoanalysis. *Behavioral Science,* 1957, **2**, 231–240.

LOEWALD, H. W. Ego and reality. *International Journal of Psychoanalysis,* 1951, **32**, 10–18.

LOEWALD, H. W. On the therapeutic action of psychoanalysis. *International Journal of Psychoanalysis,* 1960, **41**, 16–33.

RAPAPORT, D. Some metapsychological considerations concerning activity and passivity. Stockbridge, Mass.: Austen Riggs Center, 1953. (Mimeo)

RAPAPORT, D. Psychoanalysis as a developmental psychology. In B. Kaplan & H. Werner (Eds.), *Perspectives in psychological theory.* New York: International Universities Press, 1960. Pp. 209–255. (a)

RAPAPORT, D. The structure of psychoanalytic theory. *Psychological Issues,* 1960, **2**(2, Whole No. 6). (b)

WAELDER, R. *Basic theory of psychoanalysis.* New York: International Universities Press, 1960.

WHITE, R. W. Ego and reality in psychoanalytic theory. *Psychological Issues,* 1963, **3**(3, Whole No. 11).

(Received November 6, 1965)

alcoholic *Ss*, and (*b*) the speech of 10 schizophrenic *Ss*.

The 10 alcoholic *Ss* were male patients from the Alcoholic Treatment Center of the New Jersey Neuro-Psychiatric Institute. The *Ss* were diagnosed as alcoholics with no known organic involvement, ranged from 29 to 63 years in age, with a median of 37 years. Their education ranged from eighth grade to 4 years of college, with a median of twelfth grade.

The 10 schizophrenic *Ss* were chronic, regressed patients at the Clinical Investigative Unit of the Bureau of Research in Neurology and Psychiatry. Their ages ranged from 23 to 52, with a median of 36. Their education ranged from eighth grade to twelfth grade, with a median of tenth grade.

Speech was elicited from both groups in the following manner: The *S* was comfortably seated facing a mirror, with a tape recorder in view. On each occasion the interviewer used these instructions:

> I am going to ask you to tell me something about yourself. I would like you to look into this mirror and for 5 minutes say aloud whatever thoughts about yourself come to your mind. I don't care what you talk about—your appearance, what sort of person you are, how you get along with other people, how other people might think or feel about you, just so long as you talk about yourself. Are there any questions?

The interviewer left the room, returning in 5 minutes.

The taped speech was then transcribed, including all repetitions, broken words, sighs, etc. The first 100 words of the monologues were selected for cloze rating. From these speech samples, disruptions, such as broken words, were deleted; however, repetitions were retained. The speech passages were then prepared by the deletion of every fifth word, with a space of uniform length substituted. Wherever the fifth word was a proper noun or number, the next word was deleted with a corresponding number of words added to the passage, so that the last twentieth blank was always followed by four words.

The raters, assembled in one group session, received a booklet containing a face sheet with a brief set of questions on the socioeconomic variables, a sheet with numbered blanks for the "things round" test, a sheet with the 54 Kent-Rosanoff word list with blanks opposite, and four sheets of prepared speech samples, two alcoholic and two schizophrenic, which were randomly assigned regarding both order and contributor. The following instruction preceded the cloze ratings:

> This is a study of language. On each of the next four pages of your booklet is a sample of a person's speech. All punctuation has been omitted. You will notice that every so often there is a blank; every fifth word which was spoken has been left out. Your job is to fill the blank with the word you think will make the most sense. The missing words are *not* numbers or names. This is difficult and often you will have to guess. Do the best you can. It is *very important* that you fill in *all* the blanks. Work as quickly as you can.

Most raters completed the task in less than the allotted 30 minutes.

In addition to the seven socioeconomic and six verbal performance variables, a cloze score (ratio of correct to total guesses) for alcoholic speech, one for schizophrenic speech, and a total score was obtained for each rater.

For all the 279 *Ss* complete data except for the verbal association measures were available. For 143 of these *Ss* the verbal association data were also available; the others, having filled in an alternate form of the test, were not used in the present analysis.

RESULTS AND DISCUSSION

Tables 1 and 2 present the product-moment correlations and the means and standard deviations of the socioeconomic and verbal performance variables for the 279 raters with the cloze scores for alcoholics, for schizophrenics, and total cloze score. As previously noted, for the correlations with the three verbal association measures, only 143 raters were available.

It was found that ability to predict the speech of alcoholics correlates significantly[8] with only two of the socioeconomic variables, with sex of rater (.13), the females emerging as better predictors, and with being lastborn (.11). Predicting of alcoholic speech also relates to predicting schizophrenic speech (.35), as well as to one of the verbal performance measures, the SCAT score (.22).

Ability to predict the speech of schizophrenics does not correlate significantly with any of the socioeconomic variables. It does relate significantly, as has been said, to the ability to predict alcoholic speech (.35), to SCAT score (.22), and negatively to the verbal association nonantonym popular score (−.16).

Ability to predict speech, whether alcoholic or schizophrenic, relates significantly to being female (.14) and to SCAT score (.30).

As SCAT score is also significantly correlated with being female (.18), the question arises as to whether sex or rater or SCAT score is the significant factor. A partial correlation between sex of rater and cloze score with the effect of SCAT score removed shows no significant relationship between sex of rater and cloze (.09), suggesting that SCAT performance rather than sex of rater is responsible for the differential rating.

---

[8] $r = .11$ is significant at the .05 level for 279 subjects. $r = .16$ is significant at the .05 level for 143 subjects.

Of the variables studied, cloze performance in general would appear to be related significantly only to verbal ability as measured by the SCAT score, which is in line with the observations of Taylor (1957) and Rankin (1959), rather than with those of Weaver and Kingston (1963).

With regard to the differential relation of the characteristics studied to the prediction of alcoholic versus schizophrenic speech, certain somewhat suggestive findings have emerged. For instance, SCAT score is more strongly related to the prediction of alcoholic speech (.27) than to the prediction of schizophrenic speech (.22)—though the difference is not significant ($t = .72$)—suggesting that other factors may be involved in the predicting of schizophrenic speech. The negative correlation between the verbal association nonantonym popular score and the prediction of schizophrenic speech further suggests such factors. Again, it is not surprising that individuals high on popular responses would be poor at predicting schizophrenic speech. Finally, in view of the presumably more sociable nature of the alcoholic as opposed to the characteristically withdrawn schizophrenic, it is of interest that the lastborn child, presumably exposed to a different and perhaps more populated social environment, should have an advantage in guessing alcoholic, but not schizophrenic, speech.

TABLE 1

CORRELATIONS BETWEEN ALCOHOLIC, SCHIZOPHRENIC, AND TOTAL CLOZE SCORES AND SOCIOECONOMIC AND VERBAL PERFORMANCE VARIABLES

Variable	Cloze scores		
	Alcoholic	Schizophrenic	Total
Sex	.13*	.07	.14*
Birth order (firstborn)	−.04	−.03	−.04
Birth order (midborn)	−.07	.09	.00
Birth order (lastborn)	.11*	−.05	.04
Number of siblings	.00	.07	.04
Father's occupation	−.09	−.04	−.08
Father's education	−.03	−.04	−.05
Mother's education	.02	.03	.01
Social class	−.08	−.03	−.07
SCAT score	.27**	.22**	.30**
Ideational fluency—total responses	.00	.06	.03
Ideational fluency—unique responses	−.08	−.01	−.06
Verbal association—total	−.03	−.10	−.08
Verbal association—antonym populars	−.05	−.02	−.06
Verbal association—nonantonym populars	.02	−.16**	−.07
Alcoholic cloze score	1.00	.35**	
Schizophrenic cloze score	.35**	1.00	

Note.—For verbal association variables, $N = 143$; for all others, $N = 279$.
* $p < .05$.
** $p < .01$.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF SOCIOECONOMIC AND VERBAL PERFORMANCE VARIABLES

Variable	M	SD
Sex	0.74	0.44
Birth order (firstborn)	0.54	0.50
Birth order (midborn)	0.21	0.41
Birth order (lastborn)	0.24	0.43
Number of siblings	2.90	1.56
Father's occupation	3.97	1.39
Father's education	4.31	1.44
Mother's education	4.23	1.06
Social class	3.63	0.87
SCAT score	51.87	10.54
Ideational fluency—total responses	16.15	5.37
Ideational fluency—unique responses	2.61	2.25
Verbal association—total	20.96	7.56
Verbal association—antonym populars	9.63	5.71
Verbal association—nonantonym populars	11.40	3.95
Alcoholic cloze score	15.32	3.99
Schizophrenic cloze score	14.77	3.36

Note.—For verbal association variables, $N = 143$; for all others, $N = 279$.

However, while these correlations were statistically significant, they were small, limiting the generality of the results, especially as bearing on the differences between schizophrenic and alcoholic speech based on a few samples of each.

REFERENCES

AMAREL, M., & CHEEK, F. Some effects of LSD-25 on verbal communication. *Journal of Abnormal Psychology*, 1965, 70, 453–456.

CATTELL, R. B. *A guide to mental testing*. London: University of London Press, 1936.

FELDSTEIN, S., & JAFFE, J. Language predictability as a function of psychotherapeutic interaction. *Journal of Consulting Psychology*, 1963, **27**, 123–126.

FILLENBAUM, S., & JONES, L. V. An application of "cloze" technique to the study of aphasic speech. *Journal of Abnormal and Social Psychology*, 1962, **65**, 183–189.

HONIGFELD, G. Effect of hallucinogenic agent on verbal behavior. *Psychological Reports*, 1963, **13**, 383–385. (a)

HONIGFELD, G. The ability of schizophrenics to understand normal, psychotic and pseudo-psychotic speech. *Diseases of the Nervous System*, 1963, **24**, 692–694. (b)

HONIGFELD, G. Temporal effects of LSD-25 and epinephrine on verbal behavior. *Journal of Abnormal Psychology*, 1965, **70**, 303–306.

HONIGFELD, G., PLATZ, A., & GILLIS, R. D. Verbal style and personality: Authoritarianism. *Journal of Communication*, 1964, **14**, 215–218.

OSGOOD, C. E. The representational model and relevant research methods. In I. de S. Pool (Ed.), *Trends in content analysis*. Urbana, Ill.: University of Illinois, 1959. Pp. 33–88.

PLATZ, A., & HONIGFELD, G. Some effects of anxiety on the intelligibility of verbal communication in psychotherapy. *Journal of Personality and Social Psychology*, 1965, **2**, 122–125.

RANKIN, E. F., JR. The cloze procedure—Its validity and utility. In O. S. Causey & W. Eller (Eds.), *Starting and improving college reading programs*. Eighth Yearbook of the National Reading Conference. Fort Worth, Texas: Texas Christian University Press, 1959. Pp. 131–138.

RUSSEL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to the 100 words from the Kent-Rosanoff word association test. University of Minnesota, Studies on Language, **11**, 1954.

SALZINGER, K., PISONI, S., FELDMAN, R. S., & BACON, P. M. The effect of drugs on verbal behavior. Presented at the American Association for the Advancement of Science Symposium, Denver, 1961.

SALZINGER, K., PORTNOY, S., & FELDMAN, R. S. Verbal behavior of schizophrenic and normal subjects. *Annals of the New York Academy of Sciences*, **1964**.

TAYLOR, C. W. A factorial study of fluency in writing. *Psychometrika*, 1947, **12**, 239–262.

TAYLOR, W. L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, **30**, 415–433.

TAYLOR, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, 1957, **41**, 12–26.

WEAVER, W. W., & KINGSTON, A. J. A factor analysis of the cloze procedure and other measures of reading and language ability. *Journal of Communication*, 1963, **13**, 252–261.

(Received November 8, 1965)

Journal of Abnormal Psychology
1966, Vol. 71, No. 6, 449–458

EFFECTS OF ANXIETY AND DEFENSE IN A THERAPY-LIKE SITUATION[1]

FRANCES KAPLAN

Yale University

The major hypotheses were that (a) high-anxious (HA) Ss talk with less spontaneity and affect in an ambiguous condition than in a more structured one, while for low-anxious (LA) Ss the reverse relationship obtains; (b) LA Ss talk with less affect and spontaneity than HA Ss. 52 HA and LA women were each instructed to talk as a patient might in therapy. 2 sets of instructions used were intended to vary in ambiguity. HA Ss did talk with greater affect and spontaneity than LA Ss, and the predicted interaction with ambiguity occurred. The results suggested that the anxiety scale constituted a measure of defensive style. Further, the usual "HA" group is not homogeneous, but contains at least 2 subgroups.

An important aspect of a patient's ability to participate in psychotherapy is his ability to talk spontaneously and with affect about important material. Bordin (1955) suggests that facilitating the production of such material is ambiguity in such aspects of the therapeutic situation as the topics of discussion, the relationship with the therapist, and the goals the therapist has in mind for the treatment. This viewpoint implies that maximum ambiguity facilitates therapy. However, since ambiguity increases anxiety (e.g., Dibner, 1958), patients who are characteristically highly anxious might be disrupted by ambiguous conditions. Thus, Bordin hypothesizes an optimal level of ambiguity for each individual that is necessary for therapy to proceed most effectively and efficiently (also Frank, 1961). At this point, the patient is anxious enough to be motivated to talk about himself, yet not so anxious that most of his energies are consumed in self-protective efforts.

The present experiment attempts further to study the relationship between ambiguity and anxiety in a therapy-like situation. The situational manipulation employed instructions that varied in the degree of ambiguity concerning what the subject was supposed to talk about. One might question, however, whether in fact ambiguity was the most important dimension being manipulated by the instructions. Descriptively, in the "ambiguous" condition traditional free-association instructions were given, and in the "structured" the subjects were told to "talk about the things you would talk about if you were a patient who had come to a therapist." The latter, which seemed to instruct them specifically to talk about their problems, was labeled "problem focused condition."

Thus, the major interest of the present study concerns the response of high- and low-anxious subjects to free association and the problem-focused instructions. Hypothesis I is that individuals who report in a questionnaire that they experience high levels of anxiety do not talk as spontaneously or with as much affect in the free-association condition as they do in the problem-focused one. Conversely, subjects who report less anxiety talk more spontaneously and with more affect in the free-association condition than in the problem-focused one. (This hypothesis assumes that the instructions vary in degree of ambiguity and that the initial anxiety level of the low-anxious, LA, subjects is below the optimal level described by Bordin.)

Work with the anxiety scales has shown that the way subjects respond to them is significantly affected by their "defensiveness," operationally defined in these studies as "an individual's need to present himself in a favorable light" (Ruebush, Byrum, & Farnham, 1963; also I. G. Sarason, 1958). For example, the anxiety scales and the defensiveness scales are negatively correlated (e.g., I. G. Sarason, 1959a; S. B. Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960). Further, most of the existing data indicate that high- and low-defensive children have different ways of dealing with potentially threatening situations, such as school tests (Doris & Sarason, 1955; Eriksen & Davids, 1955; Sarason, Davidson, Lighthall, & Waite, 1958). Barnard, Zimbardo, and S. B. Sarason (1961) suggest that the high-defensive, LA children have difficulty utilizing their traditional modes of defense in a very permissive situation, and so tend to become anxious under these conditions. Previous work has been done exclusively with children, so the results are not necessarily directly applicable to college students or adults. The pretest data of this research suggested that the defensive behavior of LA subjects increases in the experimental situation, taking the form of interference with spontaneous talking and with affect. Keeping in mind its tentative and exploratory nature, a second hypothesis can be made. Hypothesis II is that LA subjects show more "resistance" than high-anxious (HA) subjects (S. B. Sarason[2]). Thus they do not talk as spontaneously and with as much affect as do HA subjects. ("Resistance" is operationally defined as silence, intellectual arguments against talking freely, hostility toward the experimenter, etc.)

---

[1] This paper is adapted from a dissertation submitted to Yale University in partial fulfillment of requirements for the degree of Doctor of Philosophy. The author acknowledges with many thanks the assistance of members of her committee, Seymour B. Sarason (chairman), Michael Kahn, and Norman Miller.

Hypothesis III is that all HA subjects in both conditions talk more about, either directly or indirectly, the following issues: anxiety, dependency wishes, hostility toward themselves, and hostility toward others, than LA subjects. Furthermore, HA subjects in the free-association condition talk more about them than HA subjects in the problem-focused condition (Davids & Eriksen, 1955; I. G. Sarason, 1960; I. G. Sarason & Campbell, 1962; S. B. Sarason et al., 1960).

METHOD[3]

Subjects

The Test Anxiety Questionnaire (TAQ) and General Anxiety Questionnaire (GAQ) (S. B. Sarason & Gordon, 1953; S. B. Sarason & Mandler, 1952), with the Defensiveness Scale (DS) and the Hostility Scale (HS) from I. G. Sarason's Autobiographical Survey (1959b) were given to 275 women students at Southern Connecticut State College during regular class time. Of these, 246 could be included in the study. The tests were all standardly administered by a faculty member at the school. The TAQ is a 35-item questionnaire with items about experienced anxiety and anxiety symptoms in testing situations (see S. B. Sarason et al., 1960). Mandler and Cowen (1958) report a test-retest reliability of .91 ($N = 70$) and a split-half reliability of .91 ($N = 100$). The GAQ is a 40-item questionnaire about anxiety and anxiety concomitants in a variety of nontest situations. Gordon and S. B. Sarason (1955) report a test-retest reliability of .81 ($N = 239$) over a period of 3½ months, and a correlation of .47 between the TAQ and GAQ.

The DS is composed of 23 true-false items, largely taken from the *K* scale of the MMPI (S. B. Sarason et al., 1960). The DS has been found to correlate negatively with test anxiety ($r = -.54$) and with general anxiety ($r = -.61$) when given to hospitalized neurotic and psychotic patients (I. G. Sarason, 1959b).

From the group of 246 women, the HA and LA subjects were selected randomly from the top and bottom 20% of the distribution of combined total scores on the anxiety scales. Initially an attempt was made to select half of each anxiety group from the top of the distribution on the DS and half from the lower half. Because of the negative correlation between the two scales, it was not possible to get equal groups on the DS distribution. In all, the final sample was composed of 52 women, of whom 29 were LA and 23 HA.

Procedure

The subjects were telephoned individually and asked to participate in a psychotherapy experiment as paid volunteers. When they arrived for their appointment they were met and escorted by the experimenter to a psychotherapy interview room where the sessions were held. The subject was seated in a chair directly facing a large one-way-vision mirror, and the tape recorder—located in another room—was turned on. The experimenter, who did not know the anxiety level of the subject, then gave the instructions. Approximately half of the subjects in each group received the instructions for the free-association condition, half the ones for the problem-focused condition. To subjects in both groups, the similarity of the situation to the therapeutic one was emphasized. They were asked to sit in the room and talk, and told they were being observed by a psychologist, who was behind the mirror listening to them and trying to understand them. The free-association group was then given the traditional instructions to say anything that came into their minds, without omitting anything. The problem-focused subjects were told instead to talk about "what you would feel you needed to talk about, if you were a patient who had come to a therapist."

---

[2] Personal communication, 1964.

[3] Additional information concerning the instructions and procedures, as well as tables of the raw data and more extensive descriptions of the statistical analyses are available in the original dissertation. A microfilm copy is available at the Sterling Memorial Library, Yale University, New Haven, Connecticut, under the same title as this article.

Initially there was some concern that subjects might interpret the latter instructions as indicating a role-playing situation. However, from the pretesting it became apparent that very few subjects did interpret the instructions in this manner, and the several who began with role-playing only continued with it for the first few minutes, and then went on to talk directly about themselves and their problems. Any questions asked by the subject after the instructions were given were answered by repeating part of the instructions, or by asking that the subject wait for an answer until after the experiment.

At the end of the 20-minute session, the subjects were asked to talk for 10 minutes about any dreams they had had that they remembered.[4] When this part of the experiment was completed, the experimenter talked briefly with each subject about her experience during the experiment, then asked each to rate herself on the degree of ambiguity and of anxiety she experienced during the first 20 minutes.

Dependent Variables

The main dependent variable was Bordin's Free Association Scale (FAS; 1964 [5]). Bordin (1963) reports interscorer reliabilities with five scorers ranging from .68 to .92. The scores on this measure reflect the degree of spontaneity and affect a subject shows in an interview situation, with high scores obtained by talking as a therapist might want a patient to talk in psychoanalytically oriented psychotherapy or psychoanalysis.

Several Speech Disruption (SD) measures were scored. These include (*a*) the Speech Disruption Ratio (SDR; Mahl, 1956; Mahl & Schulze, 1962), (*b*) the number of "*Ahs*," suggested by Panek and Martin (1959) as a reflection of defense against anxiety, and (*c*) the total number of words. When the SDs were scored directly from tapes, Panek and Martin (1959) reported an average reliability from four tapes of .75. Following Mahl's example (1962), every other 2-minute period of each session was scored for the SDR.

The third set of dependent variables was a modified form of the Dollard and Auld (1959) content category scoring system. This system was designed to reflect the major issues that occur in the verbal communication between a patient and a therapist. In this study, the unit used was a 2-minute interval of time instead of the sentence, since it is difficult reliably to identify sentences directly from tapes. The categories included were: anxiety, dependency, hostility, hostility toward the self, and resistance.

All of the scoring was done by the experimenter, completely blind as to subject, anxiety level, and condition. Each 10-minute unit of the session was scored independently. A second scorer then scored 11, or approximately one-quarter, of the tapes, which were randomly selected from the 43 sessions finally included in the study.

RESULTS

Subject Selection

From the initial group of 71 subjects who were called, 19 refused. A definite trend toward a higher rate of refusal by HA subjects was apparent, although not statistically significant because of the small *N*. The HA subjects who refused to come had significantly higher scores on the TAQ than the HA subjects who came and participated ($t^2 = 6.718$, $p < .025$, two-tailed). Of the 52 subjects who appeared for the experiment, nine were eliminated from most analyses because they were completely silent during one or both of the 10-minute units, and thus did not participate in a way that could be meaningfully scored on the FAS. Four of the nine were in the HA problem-focused group, three in the LA problem-focused group, and two in the LA free-association group. Although it does not reach significance, the trend was in the direction of there being a larger number of silent subjects in the problem-focused condition. There was no difference between the silent subjects and the total sample of subjects who participated in scores on the DS or on the Otis Test. All four silent subjects in the HA problem-focused group had higher absolute scores on the TAQ than any other subject in that group.

Reliability

Interscorer reliabilities were computed on the basis of 11 subjects. Each 10-minute unit of the 20-minute sessions was scored independently, so 22 Free Association Scale (FAS) scores were included. The reliability coefficient for the sum of scores on the FAS was .88. For the content categories, reliabilities ranged from .73 for resistance to .90 for dependency.[6] Every other 2-minute interval was

---

[4] This aspect of the research is peripheral to the experiment presented here, and the results are not discussed.

[5] E. S. Bordin. Free association: An experimental analogue of the psychoanalytic situation. Unpublished manuscript, 1964.

[6] The category, hostility toward self, could not be reliably scored ($r = .22$) and was therefore eliminated from further analyses.

TABLE 1

ANALYSIS BY THE METHOD OF CONTRASTS OF THE MEAN TOTAL FAS SCORES FOR EACH EXPERIMENTAL GROUP

	Free association	Problem focused	F regression	X^2 residual
HA	9.364	12.625	8.038**	5.298*
LA	9.833	7.875		

Note.—$N = 43$, $MSw = 12{,}350$.
* $p < .05$.
** $p < .005$.

scored for the Speech Disruption (SD) measures, so there were 55 independent scores (or 45 for the total number of words) for the 11 tapes. The reliability for the SDs was .89, for the total number of words .99, and for the number of *Ah*s .97.

Free Association Scale

The mean total FAS for the 43 subjects for each experimental condition is shown in Table 1. (Each subject's score is an average of her score on the first and second 10-minute unit.) Combining Hypotheses I and II yielded the a priori prediction that the highest scores would be obtained by the HA subjects in the problem-focused condition, the next highest by HA subjects in the free-association condition, then LA subjects in the free-association condition, and the lowest scores by LA subjects in the problem-focused condition. Weights of 3, 1, −1, and −3 were selected to correspond to the predicted pattern and the F obtained by a regression of the group means on the weights by the method of contrasts (Abelson, 1962). The predicted pattern did account for a significant amount of the systematic variance of the scores ($F = 8.038$, $p < .005$[7]). The residual was also significant ($\chi^2 = 5.298$, $p < .05$).

Hypothesis I, that HA subjects in the problem-focused condition would talk with more affect and spontaneity than HA subjects in the free-association condition, and that the reverse relationship would obtain for LA subjects, was further tested by a t test between the conditions for each anxiety level separately. The predicted relationship did occur between the HA subjects in the two conditions ($t^2 = 4.646$, $p < .025$), but there was no difference between the scores of the LA subjects in the two conditions ($t^2 = 1.683$, *ns*).

Hypothesis II, that LA subjects would show more resistance and therefore talk with less affect and spontaneity than HA subjects, was supported by the overall significant F for the method of contrasts. In addition, t tests comparing HA and LA subjects showed no difference between the anxiety groups in the free-association condition ($t^2 = .100$, *ns*) but significantly higher FAS scores for HA subjects in the problem-focused condition ($t^2 = 9.007$, $p < .005$). In summary, the anxiety effect was significant for the problem-focused condition but not for the free-association, while the condition effect yielded significant differences between groups of HA subjects.

Since, initially, defensiveness was thought to be an important factor in an individual's responses to the anxiety scales and the experimental situation, the FAS scores for each experimental group were computed for subjects above and below the median on the Defensiveness Scale. Although there was a slight trend for the LA high-defensive subjects to obtain lower FAS scores by talking with less affect and spontaneity than LA low-defensive subjects, this was not significant (tested by the Mann-Whitney U test).

Content Categories

For the content categories of anxiety, dependency, and hostility, the a priori hypothesis was that HA subjects in the free-association condition would obtain the highest scores, HA subjects in the problem-focused condition would receive slightly lower scores, and LA subjects in both conditions would

TABLE 2

ANALYSIS BY THE METHOD OF CONTRASTS OF THE MEAN SCORES FOR RESISTANCE FOR EACH GROUP

	Free association	Problem focused	F regression	X^2 residual
$N = 43$	HA (−2) 1.909	(−2) .750	6.489**	7.918*
	LA (2) 2.083	(2) 3.500		
$N = 52$	HA (−2) 1.909	(−2) 2.083	2.984*	10.352**
	LA (2) 2.143	(2) 3.800		

Note.—The weights are reported in parentheses beside each mean score.
* $p < .05$.
** $p < .0125$.

---

[7] All p values reported are for a one-tailed test, unless otherwise noted.

TABLE 3

COMPARISON AMONG MEANS FOR EACH EXPERIMENTAL GROUP ON THE TOTAL NUMBER OF WORDS BY THE TUKEY MULTIPLE-RANGE TEST

	$\bar{x}$	HA-FA	HA-PF	LA-FA	LA-PF
HA-Free Association	233.309	—	3.366	5.058	104.926[a]
HA-Problem Focused	236.675		—	1.692	108.292[a]
LA-Free Association	238.367			—	109.984[a]
LA-Problem Focused	128.383				—

[a] Critical value = 49.114, $p < .05$.

receive the lowest. The weights 3, 1, −2, and −2, respectively, were selected to represent this pattern, and the regression of the means on these weights determined by the method of contrasts. The only significant result was greater anxiety expressed by all HA subjects ($F = 4.324$, $p < .05$). (By an unweighted analysis of variance, again only the level effect for anxiety was significant at less than the .10 level.)

The prediction was made that LA subjects in both conditions would have higher resistance scores than HA subjects. The weights 2 (LA free association), 2 (LA problem focused), −2 (HA free association), and −2 (HA problem focused) were selected, and the F for regression computed (see Table 2). This F and the residual were both significant. Since the resistance score includes silence as one criterion, the nine subjects excluded from other analyses on the basis of their being silent were included, and the data reanalyzed. With these subjects included, the LA subjects continued to show signicantly more resistance than HA subjects. Inspection of the means (Table 2) shows that the main effect of including these silent subjects was to increase the resistance scores in the HA problem-focused group.

TABLE 4

CORRELATIONS AMONG FAS SCORES, CONTENT CATEGORIES, AND TOTAL NUMBER OF WORDS

	FAS	Anxiety	Dependency	Hostility	Resistance
FAS		.40	.20	.31	−.83
Total words	.74	.25	.11	.33	−.80

Note.—$N = 43$.

Speech Disruption Measures

There were only two findings from the SD measures; most results from analyses of variance were not significant. The HA problem-focused group had a higher *Ah* ratio (transformed to arc sines) than any other cell, and the LA problem-focused group spoke significantly fewer words than any other group (Tukey multiple-range test, see Table 3).

Intercorrelations among Variables

Intercorrelations among FAS scores, content scores, and the total number of words are shown in Table 4. The correlation between the total number of words and FAS accounts for approximately 50% of the variance. Both the FAS and the total number of words showed a small positive correlation with all of the content categories except resistance. With resistance, each had a large negative correlation.

Additional Data

Although all subjects were clearly discouraged from asking questions until the experiment was over, some did comment, question, and/or exclaim about some aspect of the situation after the experimenter gave the instructions but before she left the room. These verbalizations were scored and an analysis of variance carried out. The results

TABLE 5

MEAN TOTAL NUMBER OF QUESTIONS, COMMENTS, AND EXCLAMATIONS PRODUCED BY EACH GROUP

	Free association	Problem focused	F condition	F anxiety	F interaction
HA	1.727	2.818	1.842	7.224*	<1
LA	4.077	5.357			

Note.—$N = 49$, $MSw = 10.221$. All p values are for a two-tailed test.
* $p < .025$.

indicate that LA subjects raised significantly more total questions, comments, and exclamations than HA subjects (see Table 5). (Interscorer reliabilities for 49 tapes were .97 for the first two categories, and .75 for exclamations.)

Each subject rated herself on the degree of ambiguity or structure she perceived in the experimental situation and also the anxiety she experienced during the 20 minutes. Although an analysis of variance indicated that subjects rated the free-association condition as being significantly more ambiguous than the problem-focused condition ($F = 4.140$, $p < .05$), none of the differences reached significance at the .05 level by the Tukey multiple-range test. Inspection of the data indicates that the trend was for LA subjects in the free-association condition to rate the situation as more ambiguous than LA subjects in the problem-focused condition, while HA subjects gave equal reports to the two conditions. The analysis of variance of the self-reports on anxiety showed no significant differences.

There was no difference among the groups on the Otis IQ Test. HA subjects had significantly lower DS scores ($F = 18.239$, $p < .01$) and higher HS scores ($F = 5.741$, $p < .05$) than LA subjects. These results were expected on the basis of the correlations between the anxiety scales and the DS ($r = -.43$) and the anxiety scales and the HS ($r = .23$).

DISCUSSION AND CONCLUSIONS

In the problem-focused condition, as predicted, the LA subjects did not talk as spontaneously or with as much affect as the HA subjects. Further, there was a tendency for the LA subjects to talk with greater spontaneity and affect in the free-association condition than in the problem-focused one. Clinical observations of the subjects and of their verbalizations suggested an explanation for these results. Despite their low scores on the anxiety scales, the LA subjects in this study did not appear to be individuals who are free from the threat of experiencing significant amounts of anxiety (also Truax, 1957). Rather, on the whole, they appear to be individuals who attempt to ward off the threat of anxiety and other distressing affects by heavy reliance on the defenses of repression and/or denial. In addition to evidence of the defensive operations themselves, these subjects demonstrated many of the characteristics consistent with this style of defense, for example, tending to express and describe relatively labile and diffuse emotional experiences, etc. (Schafer, 1954). Despite the defensive efforts, at times the underlying anxiety became apparent. The following excerpt, taken from the 10-minute session on dreams, illustrates the denial as well as the underlying anxiety:

> Well, there hasn't been anything unusual about my dreams that I can remember. I've never had a nightmare. About 6, 7, 8 years ago I had a couple of bad dreams. I don't even remember what they were. Skeletons, everybody was dressed up like a skeleton. I was scared. Never had any nightmares . . . Once I had a bad . . . well, it wasn't a bad dream, it was just something that would scare you for a minute . . . I used to dream, once in a while, I'd say three or four times, that I was falling. No particular place or anything, just have a falling sensation for a couple of seconds (laughs). Then I'd wake up and it would be all right . . .

Here the anxiety is obvious—in the bad dreams—as well as the efforts to deny and minimize it.

The degree of defensive behavior evidenced by these LA subjects at the beginning of the experiment is an additional and more objective indication of the threat presented them by the situation. Despite active discouragement, these subjects asked more questions and made more comments and exclamations than HA subjects. Further, these questions were "security questions" rather than questions reflecting curiosity and interest in the process.[8] Similarly, of all the subjects run in the experiment, including approximately 15 pretest subjects, the two who refused to stay in the situation and left in an obvious panic were LA.

This view that the so-called "low anxious" subjects rely heavily on repression and denial can account for their pattern of scores. The experimental task of talking about personal problems in a simulated therapy situation was particularly threatening to them because it directly challenged their defensive position that they do not have problems, that no upsetting, disturbing, or frightening thought ever crosses their minds. Under this stress, these subjects responded with a variety of defensive maneuvers, including silence.[9]

---

[8] Category names suggested by S. Blatt, personal communication.

For this argument, it is unnecessary to attempt to decide what combination of conscious dissembling and unconscious defenses caused their low scores on the anxiety questionnaires. The speculative interpretation offered is that subjects who obtain very low scores on these scales, by whatever psychological processes, demonstrate in interview situations a number of the characteristics of the individual who relies predominantly on repression and denial, that is, of the hysteric personality. An interesting and important control for verification of this interpretation would be to induce a response bias toward concealment of pathology in subjects with varying personality styles, and see if their verbalizations would also have many hysterical features. Although they might well demonstrate several such features, such as denial of all difficulties, it seems unlikely that they would present the total personality picture described by Schafer (1954) and seen in these subjects. However, this interpretation can be offered here only as a tentative and hopefully provocative suggestion in need of more rigorous experimentation for confirmation or rejection. (The suggestion that the response set influencing the questionnaire scores reflects stable personality variables is not without precedent, for example, Couch & Kenniston, 1960; Edwards, 1957.)

The HA subjects, on the other hand, appeared to be individuals who deal with anxiety by thinking and talking about it. That is, their style of defense permits conscious awareness of anxiety-arousing ideas (also Spielberger, DeNike, & Stein, 1965). This defensive style is reflected in their higher FAS scores and the greater amount of expressed anxiety in the content, as well as other aspects of the content. The following excerpt is from the first 2 minutes from an HA subject in the problem-focused condition:

> Uh, I have a funny relationship with my family sometimes. I, uh, I never talk back, I guess that's because I've always been taught not to but . . . I'm engaged, and it seems I always blow off steam on my fiancé . . . Which isn't very fair. And uh, we've had problems. He, he is Italian, and my father thinks that he's not, not uh, lower class, but that he would be unable to move in the same circles that I would, and uh, when we first discussed marriage he said this too . . . Anthony, and uh, that was a big problem. And uh, I've tried to convince him, Anthony, that I don't feel this way, and I don't. But uh, my family are very close . . . and I, one very important thing to me is that we all be very close.

This subject was openly expressing her anxiety about a situation which was clearly important to her. Thus to ask most HA subjects to talk about themselves in this situation posed no particular difficulty, but on the contrary provided a welcomed outlet for their thoughts and concerns.

These observations on the personality characteristics of high- and low-anxious subjects suggest that with this sample of subjects the TAQ and GAQ constituted a good measure of defensive style, rather than of amount of anxiety, per se. These scales measure an individual's tolerance for the conscious experience, and acknowledgement, of anxiety, which is more specific to the experimental situation than the items on the DS originally given these subjects (also I. G. Sarason, 1957, 1960; S. B. Sarason et al., 1960).

An important group remaining to be considered is the silent subjects. Nearly half of these fell in the HA problem-focused group. Thus although some HA subjects, when asked to talk about their problems, did so spontaneously and with affect, a significant minority did not talk at all. These latter four subjects had the highest four scores on the TAQ. This suggests that they were in fact the most anxious of the HA subjects and that their anxiety was extreme enough to immobilize or disrupt them. Possibly these individuals prevented a breakdown of defensive operations by not talking. In this respect the 12 HA subjects who refused to come, and who also had higher TAQ scores than the HA subjects who came and talked, are similar to these silent subjects. Both groups deal with the threat of extreme anxiety by a form of withdrawal from the situation.

---

[9] The high correlation between the resistance, FAS scores, and the total number of words suggests that the single most important factor in all three scores is the quantity of speech.

These results suggest that at the extreme top of the distribution of anxiety-scale combined scores, there is a group of subjects who are beyond the optimal level of anxiety to talk comfortably; rather, these subjects are disrupted by their anxiety and the situation. (S. B. Sarason & Gordon, 1955, also noted a discontinuity between HA subjects in the top 14% of the distribution and those between the 14th and 28th percentage.) Frequently combined with these subjects in anxiety studies are those HA individuals who characteristically deal with anxiety by thinking and talking about it, but who are not so anxious that they withdraw from the situation, or are unduly disrupted by it. The usual LA group are individuals who deny and repress to a marked degree.

The data do not unequivocally indicate that the instructions varied mainly in the degree of ambiguity. However, there were interesting differential responses to them. As was noted above, LA subjects had somewhat more difficulty talking in the problem-focused condition, while most HA subjects seemed to perceive it as an opportunity and obligation to discuss important and troublesome concerns. A small group of HA subjects also found the problem-focused instructions potentially threatening. The fact that there was no difference between anxiety groups in the free-association condition suggests that these instructions are more neutral. By not specifically demanding anxiety-arousing content, they neither arouse as much of the defensiveness of the LA subjects, nor of the frankness and willingness to discuss personal problems of the HA subjects.

The further question of the demand characteristics of the experiment and the effects this might have had on the results must be raised. There were both implicit and explicit demands communicated to the subjects, in addition to their initial and probably diverse expectations of a "psychology experiment." However, there is no reason to believe that these factors differentially affected the experimental groups. The subject population was highly homogeneous in terms of age, socioeconomic background, education, and even exposure to academic psychology. Additional research would be needed to ascertain the specific effects of such prior attitudes and perceived demands on the subjects. The limited sample (and in particular the fact that the subjects were all women) obviously limits the generality of the results. Also, although HA subjects tend to be more compliant than LA subjects, this fact alone does not explain the results, since a significant number of HA subjects showed "noncompliance" by either refusing to come to the experiment or by refusing to talk.

Most of the results from the SD measures were not significant. A possible interpretation for the greater *Ah* ratio for the HA subjects in the problem-focused group is that the *Ah*s reflect an aspect of the defense of these subjects, comparable to the relative silence which served as a defense for the LA subjects (also Mahl & Schulze, 1962; Panek & Martin, 1959).

Contrary to the prediction, HA subjects did not express more dependency wishes and hostility than LA subjects. One interpretation is that the failure to differentiate between unconscious and conscious expression of the needs made the scoring inappropriate. (The original Dollard and Auld categories were designed to discriminate between unconscious and conscious need expression. However, in pretest data reliable discrimination was unattainable, and thus not attempted.) If the real difference between anxiety groups involves the awareness and acceptance of needs, rather than their presence or absence, this scoring would obscure the differences.

A major interest underlying this research was to provide data that has relevance for psychotherapy. Although there are many important and obvious differences between the experimental situation and the psychotherapeutic one, there are surprisingly many similarities. For example, many subjects indicated that their expectations about coming to an experiment concerning psychotherapy were closely related to fantasies that they were coming for treatment (see Footnote 5). However, since this experiment was a single interview, while therapy is an evolving process and relationship, only generalizations to the beginning phase, and possibly only the first hours of treatment are valid.

Strengthening the claim that such generalizations have validity are the results from several studies dealing directly with therapy patients in a clinic setting. Gallagher (in Snyder, 1953) and others (e.g., Frank, Gliedman, Imber, Nash, & Stone, 1957; Lorr, Katz, & Rubinstein, 1958) found that the more openly anxious the individual is, the more likely he is to remain in psychotherapy, while the more defensive, the less likely that is. In these studies, as well as in the experiment reported, the variable of acknowledged anxiety is an important factor in an individual's willingness and ability to participate in the situation. The results from this research raise the possibility that the "good" psychotherapy patient, as the "good" subject, is not necessarily more anxious than other patients, but is an individual whose style of defense permits conscious awareness of anxiety. Both the expressed and experienced anxiety and the nature of the participation in therapy are largely results of the defensive style, rather than participation being a result simply of the experienced anxiety. One category of "poor" patient, on the other hand, is the group of individuals who rely so heavily on repression and denial that they cannot permit themselves any awareness of their internal states.

Some preliminary studies on psychotherapy patients provide additional support for this interpretation. Taulbee (1958) reported that in his study the patients who continued in psychotherapy were more introspective and "more anxious, sensitive, dependent, self-doubting and have increased awareness of feelings of inadequacy, inferiority, and depression." On the other hand, those patients who terminated psychotherapy were seen as highly constricted individuals. In these patients, repression had generalized to such an extent that they were responsive to only a very limited range of emotional stimulation. Kagan (1961) found that individuals rated as introspective and considered "good" prospective therapy patients were those subjects who had conflictual thoughts available to verbal report. The nonintrospective subjects were characteristically repressors.

These studies, with the clinical impressions of the investigators, provide some support for the interpretation of LA subjects offered in this research. Their constriction through heavy reliance on denial and repression, as well as their ability to reject a task in the service of defense, results in resistance to talking spontaneously and with affect in the experimental situation, and also makes them relatively difficult patients in at least the early phase of psychotherapy.

REFERENCES

ABELSON, R. Testing a priori hypotheses in the analysis of variance. Unpublished manuscript, Yale, 1962. (Mimeo)

BARNARD, J. W., ZIMBARDO, P. G., & SARASON, S. B. Anxiety and verbal behavior in children. *Child Development*, 1961, **32**, 379–392.

BORDIN, E. S. Ambiguity as a therapeutic variable. *Journal of Consulting Psychology*, 1955, **19**, 9–15.

BORDIN, E. S. Response to the task of free association as a reflection of personality. Paper read at International Congress of Scientific Psychology, Washington, 1963.

COUCH, A., & KENNISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 1960, **60**, 151–174.

DAVIDS, A., & ERIKSEN, C. W. The relation of manifest anxiety to association productivity and intellectual attainment. *Journal of Consulting Psychology*, 1955, **19**, 219–222.

DIBNER, A. S. Ambiguity and anxiety. *Journal of Abnormal and Social Psychology*, 1958, **56**, 165–174.

DOLLARD, J., & AULD, F., JR. *Scoring human motives.* New Haven: Yale University Press, 1959.

DORIS, J., & SARASON, S. B. Test anxiety and blame assignment in a failure situation. *Journal of Abnormal and Social Psychology*, 1955, **50**, 335–338.

EDWARDS, A. E. *The social desirability variable in personality assessment and research.* New York: Dryden Press, 1937.

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *Journal of Abnormal and Social Psychology*, 1955, **50**, 135–137.

FRANK, J. *Persuasion and healing*. Baltimore: Johns Hopkins Press, 1961.

FRANK, J. D., GLIEDMAN, L. H., IMBER, S. D., NASH, E. H., JR., & STONE, A. R. Why patients leave psychotherapy. *Archives of Neurology and Psychiatry*, 1957, **77**, 283–299.

GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *Journal of Personality*, 1955, **23**, 317.

KAGAN, J. Stylistic variables in fantasy behavior: The ascription of affect states to social stimuli. In J. Kagan and G. S. Lesser (Eds.), *Contemporary issues in the Thematic Apperception Test*. Springfield, Ill.: Charles C Thomas, 1961.

LORR, M., KATZ, M. M., & RUBINSTEIN, E. A. The prediction of length of stay in psychotherapy. *Journal of Consulting Psychology,* 1958, **22**, 321–327.

MAHL, G. Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology,* 1956, **53**, 1–15.

MAHL, G., & SCHULZE, G. Psychological research in the extralinguistic area. Paper read at Interdisciplinary Work Conferences on Para-language and Kinesis, 1962.

MANDLER, G., & COWEN, J. E. Test Anxiety Questionnaire. *Journal of Consulting Psychology,* 1958, **22**, 228–229.

PANEK, D. M., & MARTIN, B. The relationship between GSR and speech disturbance in psychotherapy. *Journal of Abnormal and Social Psychology,* 1959, **58**, 402–405.

RUEBUSH, B., BYRUM, M., & FARNHAM, L. J. Problem solving as a function of children's defensiveness and parental behavior. *Journal of Abnormal and Social Psychology,* 1963, **67**, 355–362.

SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *Journal of Consulting Psychology,* 1957, **21**, 485–490.

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *Journal of Abnormal and Social Psychology,* 1958, **56**, 339–344.

SARASON, I. G. Intellectual and personality correlates of test anxiety. *Journal of Abnormal and Social Psychology,* 1959, **59**, 272–275. (a)

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *Journal of Abnormal and Social Psychology,* 1959, **59**, 37–42. (b)

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychological Bulletin,* 1960, **57**, 403–415.

SARASON, I. G., & CAMPBELL, J. P. Anxiety and the verbal conditioning of mildly hostile verbs. *Journal of Consulting Psychology,* 1962, **26**, 213–216.

SARASON, S. B., DAVIDSON, K., LIGHTHALL, F., & WAITE, R. Rorschach behavior and performance of high and low anxious children. *Child Development,* 1958, **29**, 277–285.

SARASON, S. B., DAVIDSON, K., LIGHTHALL, F., WAITE, R., & RUEBUSH, B. *Anxiety in elementary school children.* New York: Wiley, 1960.

SARASON, S. B., & GORDON, E. M. The Test Anxiety Questionnaire: Scoring norms. *Journal of Abnormal and Social Psychology,* 1953, **48**, 447–448.

SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *Journal of Abnormal and Social Psychology,* 1952, **47**, 810–817.

SCHAFER, R. *Psychoanalytic interpretations in Rorschach testing.* New York: Grune & Stratton, 1954.

SNYDER, W. V. (Ed.) *Group report of a program of research in psychotherapy.* Philadelphia: State College, 1953.

SPIELBERGER, C. D., DENIKE, L. D., & STEIN, L. S. Anxiety and verbal conditioning. *Journal of Personality and Social Psychology,* 1965, **1**, 229–239.

TAULBEE, E. S. Relation between certain personality variables and continuation in psychotherapy. *Journal of Consulting Psychology,* 1958, **22**, 83–89.

TRUAX, C. B. The repression response to implied failure as a function of the hysteria-psychasthenic index. *Journal of Abnormal and Social Psychology,* 1957, **55**, 188–193.

(Received November 15, 1965)

Prentice-Hall books in Psychology

THE SHAPING OF PERSONALITY: TEXT AND READINGS FOR A SOCIAL LEARNING VIEW by Georgia Babladelis, California State College at Hayward, and Suzanne Adams, Merrit College, California. A descriptive account of the development and functional organization of individuality, written from the point of view of social and cognitive learning theory. Includes commentary on research articles and an extensive bibliography. January 1967, approx. 536 pp., $5.95

AN INTRODUCTION TO PERSONALITY: A RESEARCH APPROACH by Donn Byrne, University of Texas. Presents a view of the field from the inside—the student is shown where ideas originate, how they are translated into research, and how theoretical speculations and research findings are interconnected. Recommended as a basic text for courses in personality and research on human behavior. 1966, 552 pp., $7.95

THE ABNORMAL PERSONALITY THROUGH LITERATURE by Alan A. Stone, M.D., McLean Hospital, Massachusetts, and Sue S. Stone. Presents a brief description of each important type of mental disorder and symptom, followed by an example of this disorder or symptom selected from literature. Each disorder is illustrated in a vivid and creative manner so that the basic concepts of mental illness and its classification can be readily learned and retained. 1966, 423 pp., $4.50

PSYCHOPATHOLOGY OF CHILDHOOD by Jane W. Kessler, Western Reserve University. Child behavior problems of all degrees of seriousness are the concern of this text. The major emphasis is on psychological factors although organic and constitutional contributions are considered in relation to specific clinical problems. ". . . an excellent text for teaching a course in the behavior problems of children."—Burton J. Michelson, Lecturer, Marquette University. 1966, 533 pp., $8.50

PERSONALITY RESEARCH: A BOOK OF READINGS Edited by Donn Byrne, University of Texas, and Marshall L. Hamilton, Washington State University. A representative picture of current research, theoretical formulations and issues essential to the science of personality—giving students direct contact with the work of some of the leading specialists in the field. 1966, 411 pp., $5.95

For approval copies, write: Box 903

Prentice-Hall, Englewood Cliffs, N. J. 07632

Oxford University Press

Perspectives in Psychopathology:

Readings in Abnormal Psychology

Edited by **James O. Palmer** *and* **Michael J. Goldstein**, University of California, Los Angeles

This collection contains twenty-eight significant clinical and research papers organized around scientific perspectives that concentrate on areas and methods of study in psychopathology rather than on disease entities. In introductions to the sections, the perspectives are closely defined and discussed. A pamphlet of questions based on the text is available to teachers. "A very useful and up-to-date collection of papers with a well-conceived organizational schema."—A. I. Rabin, Michigan State University

1966 448 pp. 44 illus. paper $4.00

Principles of Group Treatment

By **Eric Berne**, M.D., McAuley Neuropsychiatric Institute, San Francisco

This book is the first systematic treatise on the use of transactional analysis in groups. The author's study deals with questions most frequently asked by beginners in the field of psychotherapy and includes chapters on group dynamics, teaching, research and writing, and the literature of group treatment. A glossary, indices, and illustrative diagrams are included as aids to the reader.

1966 384 pp. illus. $6.50

An Introduction to Psychopathology

Second Edition

By **D. Russell Davis**, Bristol Royal Infirmary

"This work is even better in the Second Edition than it was in the first, and constitutes one of the most lucid, sensible, and eclectic treatments of the subject."—Louis Jolyon West, M.D., University of Oklahoma School of Medicine

1966 164 pp. paper $3.75

The Experience of Anxiety: *A Casebook*

By **Michael J. Goldstein** *and* **James O. Palmer**

1963 304 pp. paper $2.25

Foundations of Psychopathology

By **John C. Nemiah**, M.D., Harvard Medical School and Massachusetts General Hospital

1961 352 pp. paper $2.25

Oxford University Press / 200 Madison Ave. / New York, N.Y. 10016